THE INSULATED GATE BIPOLAR TRANSISTOR
IGBT
THEORY AND DESIGN

VISIT…

LANZAROTE
Caliente.com

IEEE Press
445 Hoes Lane
Piscataway, NJ 08854

IEEE Press Editorial Board
Stamatios V. Kartalopoulos, *Editor in Chief*

M. Akay
J. B. Anderson
R. J. Baker
J. E. Brewer
M. E. El-Hawary
R. J. Herrick
D. Kirk
R. Leonardi
M. S. Newman
M. Padgett
W. D. Reeve
S. Tewksbury
G. Zobrist

Kenneth Moore, *Director of IEEE Press*
Catherine Faduska, *Senior Acquisitions Editor*
John Griffin, *Acquisitions Editor*
Christina Kuhnen, *Associate Acquistions Editor*

THE INSULATED GATE BIPOLAR TRANSISTOR

IGBT

THEORY AND DESIGN

Vinod Kumar Khanna

IEEE

IEEE PRESS

WILEY-
INTERSCIENCE

A JOHN WILEY & SONS, INC., PUBLICATION

Copyright © 2003 by the Institute of Electrical and Electronics Engineers, Inc. All rights reserved.

Published simultaneously in Canada.

No part of this publication may be reproduced, stored in a retrieval system or transmitted in any form or by any means, electronic, mechanical, photocopying, recording, scanning or otherwise, except as permitted under Section 107 or 108 of the 1976 United States Copyright Act, without either the prior written permission of the Publisher, or authorization through payment of the appropriate per-copy fee to the Copyright Clearance Center, Inc., 222 Rosewood Drive, Danvers, MA 01923, 978-750-8400, fax 978-750-4470, or on the web at www.copyright.com. Requests to the Publisher for permission should be addressed to the Permissions Department, John Wiley & Sons, Inc., 111 River Street, Hoboken, NJ 07030, (201) 748-6011, fax (201) 748-6008, e-mail: permreq@wiley.com.

Limit of Liability/Disclaimer of Warranty: While the publisher and author have used their best efforts in preparing this book, they make no representations or warranties with respect to the accuracy or completeness of the contents of this book and specifically disclaim any implied warranties of merchantability or fitness for a particular purpose. No warranty may be created or extended by sales representatives or written sales materials. The advice and strategies contained herein may not be suitable for your situation. You should consult with a professional where appropriate. Neither the publisher nor author shall be liable for any loss of profit or any other commercial damages, including but not limited to special, incidental, consequential, or other damages.

For general information on our other products and services please contact our Customer Care Department within the U.S. at 877-762-2974, outside the U.S. at 317-572-3993 or fax 317-572-4002.

Wiley also publishes its books in a variety of electronic formats. Some content that appears in print, however, may not be available in electronic format.

Library of Congress Cataloging-in-Publication Data:

Khanna, Vinod Kumar, 1952–
The Insulated gate bipolar transistor (IGBT): theory and design / Vinod Kumar Khanna.
p. cm
"A Wiley-Interscience publication."
Includes bibliographical references and index.
ISBN 0-471-23845-7 (cloth)
1. Insulated gate bipolar transistor. I. Title.

TK971.96.B55K49 2003
621.3815′282—dc21 2003043251

Printed in the United States of America

10 9 8 7 6 5 4 3 2 1

To my daughter, Aloka, and my wife, Amita
My mother, Smt. Pushpa Khanna, and my father, Shri Amarnath Khanna

CONTENTS

Preface xv

1 Power Device Evolution and the Advent of IGBT 1

1.1 Introductory Background / 1
1.2 Insulated Gate Bipolar Transistor / 5
1.3 Advantages and Shortcomings of IGBT / 8
1.4 IGBT Structure and Fabrication / 12
1.5 Equivalent Circuit Representations / 14
1.6 Principle of Operation and Charge-Control Phenomena / 16
1.7 Circuit Modeling / 17
1.8 Packaging Options for IGBTs / 22
1.9 Handling Precautions of IGBTs / 23
1.10 IGBT Gate Driving Circuits / 23
1.11 IGBT Protection / 26
1.12 Summary and Trends / 27
Review Exercises / 28
References / 29

2 IGBT Fundamentals and Status Review 35

2.1 Device Structures / 35
2.2 Device Operational Modes / 46
2.3 Static Characteristics of IGBT / 50
2.4 Switching Behavior of IGBT / 55
2.5 Safe Operating Area (SOA) / 77
2.6 High-Temperature Operation / 83
2.7 Radiation Effects / 85

2.8 Trench-Gate IGBT and Injection-Enhanced IGBT (IEGT) / 86
2.9 Self-Clamped IGBT / 89
2.10 Ratings and Applications of IGBT / 90
2.11 Summary and Trends / 94
Review Exercises / 95
References / 97

3 **MOS Components of IGBT** **101**

3.1 General Considerations / 101
3.2 MOS Structure Analysis and Threshold Voltage / 107
3.3 Current–Voltage Characteristics of MOSFET; Transconductance and Drain Resistance / 119
3.4 ON-Resistance Model of DMOSFET and UMOSFET / 123
3.5 MOSFET Equivalent Circuit and Switching Times / 130
3.6 Safe Operating Area (SOA) / 134
3.7 Neutron and Gamma-Ray Damage Effects / 134
3.8 Thermal Behavior of MOSFET / 136
3.9 DMOSFET Cell Windows and Topological Designs / 138
3.10 Summary and Trends / 138
Review Exercises / 139
References / 140
Appendix 3.1: Derivation of Eqs. (3.2a) and (3.2b) / 141
Appendix 3.2: Derivation of Eq. (3.7) / 142
Appendix 3.3: Derivation of the Equations for Bulk Semiconductor Potential ψ_B and the Surface Charge Q_s at the Point of Transition into Strong Inversion / 146
Appendix 3.4: Derivation of Eqs. (3.33)–(3.36) / 147
Appendix 3.5: Derivation of Eq. (3.39) / 150
Appendix 3.6: Derivation of Eq. (3.49) / 152

4 **Bipolar Components of IGBT** **155**

4.1 PN Junction Diode / 155
4.2 P-I-N Rectifier / 172
4.3 Bipolar Junction Transistor / 180

4.4 Thyristor / 189
4.5 Junction Field-Effect Transistor (JFET) / 198
4.6 Summarizing Remarks / 198
Review Exercises / 198
References / 200
Appendix 4.1 Drift and Diffusion Current Densities / 201
Appendix 4.2 Einstein's Equation / 204
Appendix 4.3 Continuity Equation and Its Solution / 206
Appendix 4.4 Solution of the Continuity Equation (4.41) / 209
Appendix 4.5 Derivation of Eq. (4.50) / 212
Appendix 4.6 Derivation of Current Density Equations (4.55) and (4.56) / 217
Appendix 4.7 Transistor Terminal Currents [Eqs. (4.57) and (4.58)] / 220
Appendix 4.8 Common-Base Current Gain α_T [Eq. (4.63)] / 225

5 Physics and Modeling of IGBT **229**

5.1 PIN Rectifier–DMOSFET Model of IGBT / 230
5.2 Bipolar Transistor–DMOSFET Model of IGBT by Extension of PIN Rectifier–DMOSFET Model / 241
5.3 Bipolar Transistor–DMOSFET Model of IGBT with Device–Circuit Interactions / 251
5.4 Concluding Comments / 280
Review Exercises / 281
References / 283
Appendix 5.1 Solution of Eq. (5.8) / 285
Appendix 5.2 Derivation of Eqs. (5.33) and (5.34) / 286
Appendix 5.3 Derivation of Eq. (5.35) / 288
Appendix 5.4 Derivation of Eq. (5.38) [Solution of Eq. (5.35)] / 289
Appendix 5.5 Derivation of Eqs. (5.40)–(5.42) / 290
Appendix 5.6 Derivation of Eq. (5.44) / 293
Appendix 5.7 Derivation of Eqs. (5.81) and Construction of Equivalent Conductive Network for 1-D Linear Element / 297

6 Latchup of Parasitic Thyristor in IGBT 303

6.1 Introduction / 303
6.2 Static Latching / 306
6.3 Dynamic Latching / 308
6.4 Latching Prevention Measures / 317
6.5 Latching Current Density of Trench-Gate IGBT / 340
6.6 Summarizing Remarks / 342
Review Exercises / 342
References / 344
Appendix 6.1 Equation (6.15) / 345
Appendix 6.2 Equation (6.20) / 347

7 Design Considerations of IGBT Unit Cell 349

7.1 Semiconductor Selection and Vertical Structure Design / 349
7.2 IGBT Design by Analytical Calculations and Numerical Simulations / 362
7.3 Optimization of N-Buffer Layer Structure / 381
7.4 Field Ring and Field Plate Termination Design / 385
7.5 Surface Ion-Implanted Edge Termination / 394
7.6 Reduced Surface Electric Field (RESURF) Concept for Breakdown Voltage Enhancement in Lateral IGBT / 395
7.7 Concluding Comments / 398
Review Exercises / 399
References / 400
Appendix 7.1 Multiplication Coefficient M / 402
Appendix 7.2 V_{BR} Equation / 403
Appendix 7.3 Avalanche Breakdown Voltage (V_{B}) / 404
Appendix 7.4 Punchthrough Voltage (V_{PT}) / 406
Appendix 7.5 $BV_{\mathrm{CYL}}/BV_{\mathrm{PP}}$ Equation / 406

8 IGBT Process Design and Fabrication Technology 411

8.1 Process Sequence Definition / 412
8.2 Unit Process Steps / 428
8.3 Process Integration and Simulation / 443

Review Exercises / 450
References / 452
Appendix 8.1 Thermal Oxidation of Silicon / 453
Appendix 8.2 Derivation of Eqs. (8.3)–(8.5) / 458

9 Power IGBT Modules 465

9.1 Paralleling IGBTs, and Integration of Logic Circuits with Power Components / 465
9.2 Power Module Technologies / 470
9.3 Isolation Techniques / 475
9.4 Integrable Devices: Bipolar, CMOS, DMOS (BCD), and IGBT / 478
9.5 Power IGBT Driving, Temperature Sensing, and Protection / 479
9.6 Parasitic Components of IGBT Module Package / 482
9.7 Flat-Packaged IGBT Modules / 484
9.8 Desirable Features and Reliability of IGBT Modules / 486
9.9 Module Heat Sinks and Cooling / 489
9.10 Material Requirements for High-Power IGBT Modules / 490
9.11 State-of-the-Art and Trends / 491
Review Exercises / 495
References / 496

10 Novel IGBT Design Concepts, Structural Innovations, and Emerging Technologies 499

10.1 Trade-Off Between ON-State Voltage Drop and Switching Losses / 499
10.2 Parallel and Coupled PIN Diode–PNP Transistor Model of Carrier Distribution in the ON State of Trench IGBT / 502
10.3 Non-Self-Aligned Trench IGBT for Superior ON-State Performance / 505
10.4 Dynamic N-Buffer Insulated Gate Bipolar Transistor (DB-IGBT) / 505
10.5 Lateral IGBT with Reverse Blocking Capability / 509

10.6 Lateral IGBT with High-Temperature Latchup Immunity / 511
10.7 Self-Aligned Sidewall-Implanted N^+-Emitter Lateral IGBT (Si-LIGBT) with High Latchup Current Capability / 512
10.8 Improved LIGBT Structure for Larger FBSOA / 513
10.9 Lateral IGBT with Integrated Current Sensor / 514
10.10 Dielectrically Isolated Fast LIGBTs / 515
10.11 Lateral IGBT in Thin Silicon-on-Insulator (SOI) Substrate / 516
10.12 Lateral Trench-Gate Bipolar Transistor (LTGBT) for Improved Latchup Characteristics / 517
10.13 Trench Planar Insulated Gate Bipolar Transistor (TPIGBT) / 518
10.14 Clustered Insulated Gate Bipolar Transistor in Homogeneous Base Technology (HB-CIGBT) / 520
10.15 Trench Clustered Insulated Gate Bipolar Transistor (TCIGBT) / 522
10.16 Double-Gate Injection-Enhanced Gate Transistor (DG-IEGT) / 523
10.17 SiC IGBTs / 525
10.18 Summary and Trends / 527
Review Exercises / 527
References / 529
Appendix 10.1 Electron Current at the Collector Junction / 530
Appendix 10.2 Transient Base Stored Charge $Q_{\text{base}}(t)$ / 531
Appendix 10.3 Depletion Width in the Presence of Mobile Carrier Concentration / 532
Appendix 10.4 Modulated Base Resistance (R_b) / 533
Appendix 10.5 ON-State Voltage Drop Due to Recombination in the End Regions of the PIN Diode in the IGBT / 534
Appendix 10.6 Energy Loss / 535
Appendix 10.7 Excess Carrier Concentration p_w in the N^- Base of TIGBT at the Emitter End / 536
Appendix 10.8 ON-State Voltage Drop Across the N^- Base of IGBT / 542

11 IGBT Circuit Applications 545

11.1 DC-to-DC Conversion / 545
11.2 DC-to-AC Inversion / 558

11.3 AC-to-DC Conversion / 570
11.4 Soft-Switching Converters / 575
11.5 IGBT Circuit Simulation / 587
11.6 Applications of IGBT Converters / 592
11.7 Summarizing Remarks / 604
Review Exercises / 605
References / 607

Index 609

About the Author 627

PREFACE

The insulated gate bipolar transistor (IGBT) represents the most commercially advanced device of a new family of power semiconductor devices synergizing high-input impedance MOS-gate control with low forward-voltage drop bipolar current conduction. It reduces the size and complexity of controlling circuitry, thereby drastically reducing the system cost. Today, it is finding widespread applications in the medium-power and medium-frequency range in uninterruptible power supplies, industrial motor drives, and domestic and automotive electronics. During recent years, no other single device has been able to revolutionize the power device scenario and cast its impact on life of the common man as much as IGBT alone has done as a power conditioning device in domestic, consumer, and industrial sectors. Power is the *life blood* driving all electrical installations, machines, trains, computers, telecommunication networks, entertainment, and other household equipment, all over the world.

Despite the growing interest in this device since its conception, no book is currently available which is, to the best of my knowledge, completely devoted to the physics and technology of IGBT. There is a dearth of generalized treatises on physics and technology of semiconductor devices. Presently available books deal with semiconductor device physics, power semiconductor devices, thyristor physics, field-effect and bipolar transistor physics, MOS physics, and related device technologies. The overwhelming pervasion of IGBT in industrial and consumer electronics warranted publication of a new book that comprehensively treats the subject. The enormous interest in IGBT constituted my principal motivation in undertaking the project of writing this book.

As its title indicates, this book has a singularity of focus on IGBT. However, it goes without saying that IGBT represents an interesting combination of PIN diode, bipolar transistor, bipolar thyristor, and power DMOSFET properties. So this text on IGBT prepares the reader not only with regard to IGBT but also with regard to the aforesaid devices that work in harmony resulting in IGBT characteristics. The expansive topical coverage of this book therefore incorporates useful material from both MOS and bipolar aspects, greatly enhancing the utility of the book. To elaborate, the forward conduction characteristics of IGBT are controlled by conductivity modulation of the PIN diode as well as MOSFET channel length. The latching of the IGBT is governed by the current gain of the bipolar transistor and regenerative thyristor action. A smaller channel length yields low forward drop device but makes it vulnerable to latching of built-in parasitic bipolar thyristor. Controlling the forward blocking capability of the IGBT requires careful attention to planar floating field ring termination design. The reverse blocking voltage is dictated by beveling during chip dicing. Likewise, the turn-off time of IGBT is determined by the carrier lifetime and hence the reverse recovery waveform of the PIN diode. So, if we look with this broad perspective, learning about IGBT requires a good knowledge about these constituent devices. Thus although the focus is on a singular device, the remaining devices are automatically a part of the overall picture. The era of MOS-bipolar combination devices has already dawned, and the book seeks to introduce the reader to this new era of intermixed technologies.

This book is written at the tutorial level to fulfill the needs of power device courses in electrical and electronics engineering and microelectronics engineering. The targeted audience of this book also includes practicing engineers and scientists. The students of today are the professionals of tomorrow. A careful blending of a tutorial design for students and specialist design for the practitioners has been made. By providing a large number of examples sprinkled throughout the text, as well as appending both questions and problems at the end of each chapter, it is hoped that classroom adaptation of the book will be easy with proper selection of course material. Up-to-date end-of-chapter references will provide the researcher a useful guide to the literature on IGBT. Thus the book caters to the requirements of a wide cross section of readership embracing students, professionals, and researchers.

A comprehensive, in-depth, and state-of-the-art treatment of the subject has been provided, encompassing a wide range of topics. Chapter 1 introduces the reader to the power semiconductor device scenario, the need for MOS-bipolar combination devices, and the birth of the IGBT. The working principle of IGBT is described in a simple way. The IGBT equivalent circuit is introduced, and the SPICE model is discussed. Packaging and handling precautions of IGBTs, gate driving circuits, and protection techniques are briefly presented.

Chapter 2 summarizes the basic types of IGBTs, their operational features, performance characteristics, limitations, specifications, and applications. Lateral and vertical IGBT structures are discussed. Nonpunchthrough and punchthrough types of IGBTs are explained. Their doping profiles and operational differences are described. Different modes of operation of IGBTs, such as forward conduction and blocking modes, are dealt with. IGBT turn-on and turn-off with resistive and inductive loads are analyzed. Soft-switching concepts are outlined. Effects of temperature and nuclear irradiation on IGBT characteristics are pointed out. The working of trench-gate and self-clamped IGBTs is addressed.

Chapter 3 covers the fundamentals of MOS structure including thermal equilibrium energy-band diagram, flat-band voltage, threshold voltage, capacitance effects, power DMOSFET structures, ON-resistance components, safe operating area, radiation and thermal effects on device characteristics, DMOSFET geometrical topologies, and so on, which are essential for understanding the physical principles of operation of IGBT.

Chapter 4 presents the theory of bipolar devices such as the PN-junction diode, the PIN rectifier, the bipolar junction transistor, the thyristor, and the junction field-effect transistor. After perusal of this chapter, the reader will be able to understand the essential principles of bipolar device operation.

From Chapter 5 onwards, the study of IGBT models begins, including static, dynamic, and electrothermal behavior. Discussions of PIN rectifier–DMOSFET and bipolar transistor–DMOSFET models of IGBT are followed by analytical models of ON-state carrier distribution, two-dimensional effects, modeling of device–circuit interactions, transient analysis of IGBT circuits, and so forth.

Because latching is a serious problem with IGBTs, this issue is discussed in detail in Chapter 6, outlining the causes of latching and the techniques of providing latching immunization of IGBT structure. After explaining static and dynamic latchup, methods of latchup prevention are dealt with exhaustively.

Since the IGBT is a conglomeration of millions of elementary cells, Chapter 7 delves into the design techniques of IGBT unit cell using computer-aided design tools. The discussion begins with semiconductor selection and vertical structure design; followed by emitter and base doping profiles and channel length, transconductance and forward voltage drop, trade-off between conduction and switching losses, unit cell layout design, and intercell spacing; then N-buffer layer structural optimization; and concludes with field ring and field plate termination design, as well as other techniques of junction edge termination for surface electric field minimization and breakdown voltage enhancement.

Chapter 8 describes the enabling technologies for power IGBT fabrication. Each unit process step is discussed, and the various steps are integrated for IGBT realization. Main steps are starting silicon preparation, epitaxial growth, polySi deposition, gate oxide fabrication, diffusion and ion implantation, mask making and microlithography, dry etching, and plasma processes, trench excavation, metallization, encapsulation, and electron irradiation for lifetime tailoring. Process simulation is also reviewed.

Chapter 9 addresses the subject of power IGBT modules and the associated technologies. Discussion of logic circuits and power device integration is followed by a summary of isolation techniques. Different types of protection and other accessories in the module are described. Flat-pack modules and materials technology for modules are also examined.

Chapter 10 provides both a retrospective as well as a bird's-eye view of futuristic IGBT technologies. It surveys new design ideas and IGBT structures, giving projections on future trends in this rapidly expanding field. Structures considered include the non-self-aligned trench IGBT, dynamic N-buffer IGBT, lateral IGBTs with reverse blocking capability and high-temperature latch-up immunity, self-aligned sidewall-implanted N^+ emitter lateral IGBT with high latchup current capability, LIGBT structure for larger FBSOA, lateral IGBT with integrated current sensor, dielectrically isolated fast LIGBT, lateral IGBT in thin (SOI) substrate, lateral trench-gate bipolar transistor, trench planar IGBT, clustered IGBT in homogeneous base technology, trench-clustered IGBT, double-gate injection-enhanced gate transistor, and many others.

Finally, Chapter 11 gives a perspective of the proliferating applications of IGBTs in circuits such as motor drives, automotive ignition, power supplies, welding, induction heating, and so on. Different types of converters such as DC-to-DC converters, DC-to-AC converters and AC-to-DC converters are mathematically analyzed. Soft-switching converters are touched upon. SABER and SPICE circuit models and design methods are also discussed.

It is earnestly hoped that the above topical coverage of this book will be useful for graduate/postgraduate students and researchers in this field. The book will serve as a textbook cum reference book on the subject. If the book serves the purpose of those for whom it is intended, I will deem my endeavors amply rewarded.

Although utmost care has been taken to ensure accuracy in presentation and content, no work can claim to be error-free and complete. Suggestions for improvement are cordially welcomed from our readers.

ACKNOWLEDGMENTS

It gives me immense pleasure to thank the director, senior scientists, and my colleagues at CEERI, Pilani, for encouragement in my efforts. I wish to thank the group leader and members of the erstwhile power device group, with whom I shared many insights into the power device physics and technology over our years of working together. I am obliged to Prof. Dr. Arnold Kostka and Mr. B. Maj, Technical University, Darmstadt, for guiding me into the simulation field.

It was a great opportunity working with my editor, Ms. Christina Kuhnen. I appreciate her prompt response and quick action on my problems, without which it would not have been possible to adhere to the schedule. I am grateful to the reviewers of the manuscript for their constructive criticism and for pointing out many errors and omissions, thereby bringing the book to its present form and contents.

Any new book owes its origin to its predecessors as well as to research papers, reports, and review articles. I am indebted to the numerous authors of these works, many of whom are listed in the reference section at the end of each chapter. The interested reader is referred to the excellent works cited in the references to acquire an indepth knowledge of any specialized topic.

I am indebted to Dr. P. K. Khanna and Mr. Vijay Khanna for moral support.

Finally, I thank my daughter and wife for their love, patience, and understanding and for tolerating my long hours of work with grace, during the course of this project spanning over two years. Thanks from my heart to all of the above and also to anyone who may have directly or indirectly helped me in this work and whom I may have forgotten to mention.

VINOD KUMAR KHANNA

Pilani, India
June 2003

The Insulated Gate Bipolar Transistor IGBT Theory and Desig, by Vinod Kumar Khanna
ISBN: 0-471-23845-7 © 2003 by the Institute of Electrical and Electronics Engineers

ABOUT THE AUTHOR

Vinod Kumar Khanna, was born in Lucknow, India in 1952. He is currently a senior scientist in the Solid-State Devices Division of Central Electronics Engineering Research Institute in Pilani, India. He received his Ph.D. in Physics in 1988 from Kurukshetra University and during the last 23 years has been involved extensively in device, process design, and fabrication of power semiconductor devices. His work has focused most notably on high-current and high-voltage rectifiers, high-voltage TV deflection transistors, the power Darlington transistor, inverter grade thyristor, and power DMOSFET and IGBT.

Dr. Khanna has published more than 30 research papers in international journals and conference proceedings and authored 2 books. He has presented papers at the IEEE-IAS Annual Meeting in Denver, Colorado in 1986 and worked as a guest scientist at Technical University, Darmstadt, Germany in 1999. He is a fellow of the IETE in India and is a life member of the Semiconductor Society and Indian Physics Association.

1

POWER DEVICE EVOLUTION AND THE ADVENT OF IGBT

1.1 INTRODUCTORY BACKGROUND

Power semiconductor devices are the essential components determining the efficiency, size, and cost of electronic systems for energy conditioning. The proliferating demand of controllable power electronic systems has promoted research on novel device materials, structures, and circuit topologies [1–7]. Present-day power devices are invariably fabricated using silicon as the base material. Among the upcoming semiconductor materials, silicon carbide has attracted the most attention [8–10]. The higher breakdown field of SiC (2.2×10^6 V/cm for 4H-SiC, 2.5×10^5 V/cm for Si) enables it to offer a projected 200-fold reduction in specific ON resistance as compared to Si devices. SiC devices also promise superior high-temperature performance due to the large energy gap (4H-SiC: 3.26 eV; Si: 1.12 eV), high thermal conductivity (4H-SiC: 4.9 W/cm; Si: 1.5 W/cm), high chemical inertness, high pressure, and radiation resistance of this material.

Power device and process design engineers worldwide are relentlessly searching for the *perfect semiconductor switch* defined by the following characteristic features: (i) *Very low driving losses*: The switch has high input impedance so that the drive current is infinitesimally small. Furthermore, the drive circuit is simple and inexpensive. (ii) *Insignificant ON state or forward conduction losses*: The forward voltage drop at the operating current is zero. Additionally, the operational current density is large, making the chip small in size and cost-effective for a given current-carrying capability. (iii) *Minimal*

The Insulated Gate Bipolar Transistor (IGBT) Theory and Design, By Vinod Kumar Khanna.
ISBN 0-471-23845-7 © 2003 Institute of Electrical and Electronics Engineers

OFF state or reverse blocking losses: Infinitely large reverse blocking voltage together with zero leakage current, even when exposed to elevated temperatures. (iv) *Extremely low switching losses*: Both the turn-on and turn-off times approach zero. In direct current (time period = ∞) and low-frequency (large but finite time period) applications, these losses are very small because the switching times are much less than the respective periodic times.

Advancements in power devices have revolutionized power electronics, and today's market offers a wide spectrum of devices intended for different applications. In applications where *gate turn-off capability is not necessary, thyristors or silicon-controlled rectifiers* (SCRs), the highest power density devices, have been the workhorse of power electronics [5–6], carrying high forward currents of ~ 3500 A with a forward drop < 2 V and withstanding ≥ 6000 V in the reverse direction. Thyristors have long been the solo devices catering to the megawatt power range, available in ratings like 12 kV/1.5 kA, 7.5 kV/1.65 kA, 6.5 kV/2.65 kA, and so on. They are classified as: *phase-control thyristors* used for a 50/60-Hz AC mains line and the *inverter thyristors* for higher frequencies of ~ 400 Hz. Typical turn-on and turn-off times are 1 and 200 μsec. Thyristors are widely used in high-voltage DC (HVDC) conversion, static var compensators, solid-state circuit breakers, large power supplies for electrochemical plants, industrial heating, lighting and welding control, DC motor drives, and so on.

As turn-off is accomplished by collector–emitter voltage reversal in conventional thyristors, in applications where the *load current is both turned on and turned off by the input signal, power bipolar junction transistors* (BJTs) have been extensively used. *Modular double or triple Darlington transistors* (1200 V, 800 A) are used in converters with switching frequencies up to several kilohertz. Although bipolar transistors have turn-off time < 1 μsec, they need very high base current drive both in the ON state and during turn-off. A competing device is the gate turn-off thyristor (GTO). It has forward current capability much higher than the BJT but requires excessively high gate drive current (750 A for 4000 V, 3000 A GTO). Its switching frequency is limited to 1–2 kHz with $t_{\mathrm{on}} = 4$ μsec and $t_{\mathrm{off}} = 10$ μsec. GTOs are used in DC and AC motor drives, uninterruptible power supply (UPS) systems, static var compensators, and photovoltaic and fuel cell converters from a few kilowatts to several megawatts of power. Improvements in the GTO structure, the gate drive, the packaging, and the inverse diode led to a new switching component, the *integrated gate commutated thyristor* (IGCT), a hard-switched GTO, which may be viewed as the hybridization of a modified GTO structure with very low inductive gate drive. Also, 4.5-kV and 5.5-kV IGCTs with currents up to 4 kA, as well as 6-kV/6-kA IGCTs [11–12], are commercially available, with further possibility of extension up to 10 kV, depending on market demand.

Another device accepted for *gate turn-off applications* is the *vertical double-diffused MOSFET* (VDMOSFET) [13]. Its gate drive current is very low, and 500-V/50-A VDMOSFET devices have switching frequencies of ~ 100 kHz with turn-on and turn-off times below 100 nsec. Fast switching speed, ease of drive, wide safe operating area (SOA), and capability to withstand high rates of rise of ON-state voltage (dV/dt) have made VDMOSFETs the logical choice in power circuit designs. However, VDMOSFETs operate by unipolar conduction. So, their ON resistance drastically increases with drain-source voltage capability, restricting their exploitation to voltages less than a few hundred volts. Moreover, as the voltage rating increases, the inherent reverse diode shows increasing reverse recovery

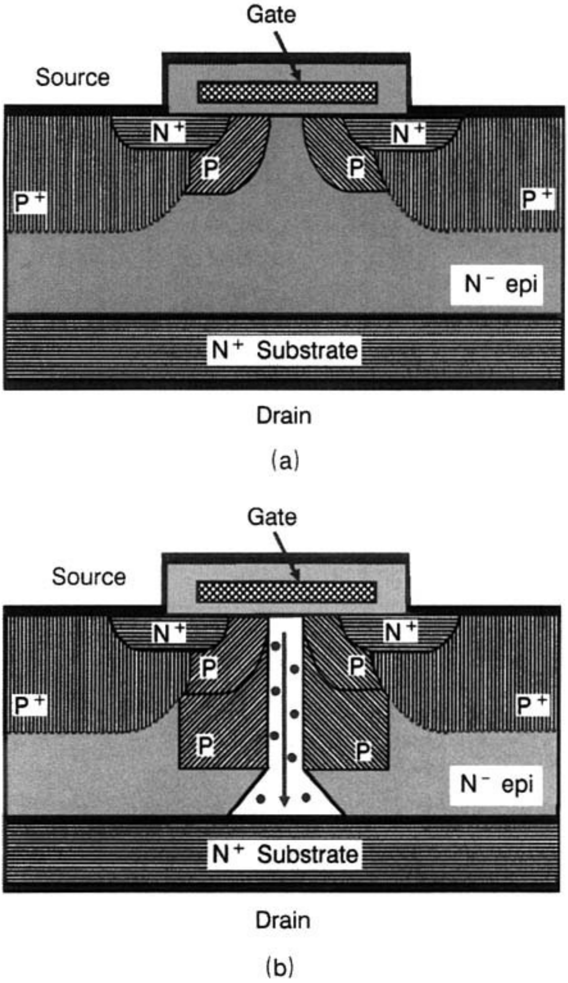

Gate
Source
N+
N+
P
P
P+
P+
N- epi
N+ Substrate
Drain
(a)
Gate
Source
N+
N+
P
P
P
P
P+
P+
N- epi
N+ Substrate
Drain
(b)

Figure 1.1 Structure of (a) conventional MOSFET and (b) Cool MOSFET.

charge (Q_{rr}) and reverse recovery time (t_{rr}), causing more switching losses. Power VDMOSFETs have gained a strong foothold in low-voltage, low-power, and high-frequency applications such as switch-mode power supplies (SMPS), brushless DC motor (BLDM) drives, solid-state DC relays, automobile power systems, and so on. A new approach to reduce the high-voltage-sustaining drift zone resistance is offered by the Cool-MOS concept [14–15], allowing

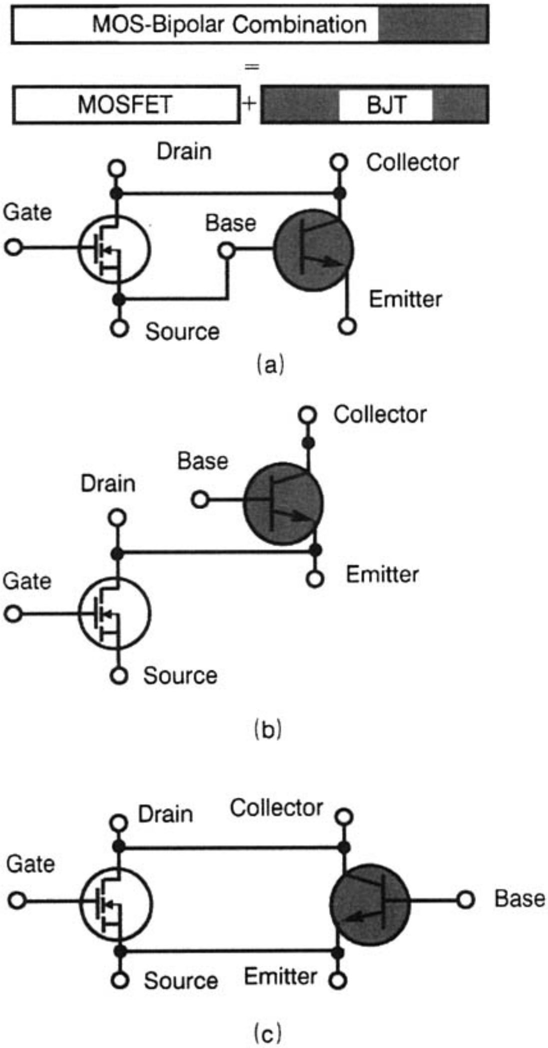

MOS-Bipolar Combination
=
MOSFET
+
BJT
Drain
Collector
Gate
Base
Emitter
Source
(a)
Collector
Base
Drain
Gate
Emitter
Source
(b)
Drain
Collector
Gate
Base
Source
Emitter
(c)

Figure 1.2 MOS-bipolar combinations. (a) Darlington configuration. (b) Series or cascade configuration. (c) Parallel or cascode configuration.

reduction of ON resistance by a factor of 5–10 compared to conventional MOSFETs having equal area, in the breakdown voltage range 600–1000 V. Here vertical P stripes are inserted into the N-drift zone (Fig. 1.1). Due to the finely structured sequence of opposite polarity layers, a marked increase in doping occurs in this zone. In the blocking state, with increasing drain–source voltages, the space-charge region at the border between P and N stripes expands, eventually leading to the depletion of the epilayer. The OFF-state voltage therefore comprises both horizontal and vertical components. Due to horizontal extension of the depletion region, drift region thickness need not be large, leading to lower conduction and switching losses and also requiring less gate drive power. For withstanding higher voltages, the area with P stripes is made larger. Reduction in doping is not necessary, as in conventional MOSFETs. Thus in Cool MOSFET, an extra P-doped region is introduced in the N-drift region. This allows a much higher breakdown voltage to be achieved using a much higher doping concentration for the N-drift region than in a conventional MOSFET. The use of a high doping concentration for the N-drift region reduces the ON resistance of the device.

Thus we find that amongst the presently available power switches, each offers distinct advantages in certain applications but suffer from shortcomings in other areas. Thus, it was considered worthwhile to blend the properties of MOSFET and bipolar devices. Indeed, the amelioration of device parameters toward the ideal switch was considerably accelerated by the idea of *MOS–bipolar combination.* In the beginning, many MOS–bipolar merger alternatives were explored. The performance characteristics and limitations of the chief combinations are pointed out below. The *Darlington configuration* (Fig. 1.2a) provides a high current gain at high output currents but gives a larger forward voltage drop than a single transistor and longer turn-off time because negative base drive cannot be applied to the BJT base during turn-off. Consequently, it exhibits high switching losses. In the *series or cascade configuration* (Fig. 1.2b) the drawbacks include the increase of forward drop and the need to drive one gate along with one base. In the *parallel or cascode Configuration* (Fig. 1.2c), the BJT must be driven in harmony with the MOSFET for turn-off loss minimisation, thus restricting the useful cut-off frequency. The breakthrough overcoming the above limitations was achieved with the success of the insulated gate bipolar transistor (IGBT).

1.2 INSULATED GATE BIPOLAR TRANSISTOR

Other names of this device include the insulated gate rectifier (acronym IGR), conductivity-modulated FET (COMFET), gain-enhanced MOSFET (GEMFET), BiFET (bipolar FET), and injector FET. It is a prime member of the family of MOS–bipolar combination devices. Other members of this family are the MOS-gated thyristor (MOS-SCR) and MOS-controlled thyristor (MCT).

The IGBT was first demonstrated by Baliga in 1979 [16] and then in 1980 by Plummer and Scharf [17], by Leipold et al. [18], and by Tihanyi [19]. Advantages of IGBT were comprehensively described by Becke and Wheatley [20] and by Baliga et al. in 1982 [21] and 1983 [22]. More work was carried out by Russell [23], Chang et al. [24], Goodman et al. [25], Baliga et al. [26], Yilmaz et al. [27] and Nikagawa et al. [28]. The IGBT was commercially introduced in the marketplace in 1983. Since then there has been a significant improvement in the device ratings and characteristics, from the initial 5 kW for discrete IGBTs to more than 200 kW for IGBT power modules. Presently, several large companies are manufacturing this device, notable among them being IXYS Corporation, International Rectifier, Powerex, Philips, Motorola, Fuji Electric, Mitshubishi Electric, Hitachi, Toshiba, Siemens, Eupec, and so on. Today, the IGBT is an established replacement of the power BJT, Darlington transistor, MOSFET, and GTO thyristor in the medium voltage (600–2500 V), medium power (10 kW), and medium frequency range up to 20 kHz. Also, 600-V/50-A IGBTs capable of hard switching at 150 kHz are commercially available. Just as the power MOSFET has replaced the BJT in low-voltage applications (< 200 V), the IGBT has replaced the BJT in the medium-voltage range 200–2000 V and is suitable for compact smart power modules. Modules with 6500-V blocking voltage capability and 200-, 400-, and 600-A current have been reported. High-voltage IGBTs are used for electric traction such as streetcars and locomotives. High-power IGBTs are challenging the dominance of GTOs in the megawatt range due to their high speed, large RBSOA, and easy controllability. However, the available power ratings of IGBTs are lower than those of GTOs, up to a rated switch power of 36 MVA (6 kV, 6 kA). MOS-SCR and MCT are alternative candidates for these applications. The injection-enhanced IGBT or IEGT [29] is a promising candidate as a next-generation high-power MOS-gated device, which can replace GTO. Basically, IGBTs operate like a bipolar transistor and have a smaller carrier accumulation in the N-type high resistance layer. So, IGBTs with forward blocking voltage > 1700 V suffer from a much larger ON-state voltage drop than do gate turn-off thyristors (GTOs). To reduce the ON-state voltage, a carrier profile similar to the GTO is adopted in the IEGT, retaining the easy gate drivability and turn-off capability of IGBT. The 4500-V IEGT has a forward drop V_F of 2.5 V at 100 A/cm^2. Current density of IEGT at $V_F = 2.5$ V is 10 times that of UMOS–IGBT (U-groove metal-oxide semiconductor–IGBT).

Today, minimum feature sizes in IGBT chips are shrinking down to 1 μm and submicron technologies using direct stepping on wafer. The IGBT is the most widely used power device in the medium power and medium frequency range, finding widespread applications in AC motor drives, traction control, inductive heating systems, radiological systems (X-ray tubes), uninterruptible power supplies (UPS), switch-mode power supplies (SMPS), static var and

Table 1.1 Applications of Different Power Electronic Sectors

Sector No.	Power Electronics Sector	Applications
1	Low-power sector (< 10 kW)	Switching power supplies for computers, printers, facsimile machines, and consumer electronics; automotive electronics, heating and lighting circuits, small motor drives, and UPS.
2	Mid-power sector (between 10 kW and 1 MW)	Solid-state drives for multi-horsepower induction motors, UPS, and machines for factory automation (using smart power ICs and modules); heating, ventilation, and air-conditioning equipment.
3	High-power or megawatt-power sector	Solid-state motor drives for heavy motors, HVDC, UPS, etc.

harmonic compensators, and so on. Table 1.1 shows that major marketing opportunities for power electronics reside in the low- and medium-power sectors. As the IGBT pervades these sectors, the utility of this device escalates.

A close look at the power electronics scenario reveals that the area between 100 V and 1000 V has vastly benefited from IGBT development and modular packaging concepts. IGBTs have gained immense importance since their introduction in the market in 1983. The IGBT is readily interconnected with control circuitry in low-cost plastic modules that are used for driving small machines for factory automation. The high output impedance of IGBTs allows parallel connection of many IGBTs. No device draws more current than its neighbors, resulting in better current sharing. So for higher load current applications, current up-scaling is accomplished by paralleling several devices. Today, 600-V, 1200-V, 2500-V, and 3300-V IGBTs/IGBT modules are commercially available up to current ratings of 2400 A. Also, 4.5- and 6.5-kV IGBT modules have been reported.

High-current and high-voltage IGBTs (> 1700 V, 1000 A) are used for traction and industrial applications. Both IGBTs and IGCTs have the potential to decrease the cost and increase the power density of pulse-width modulation voltage-source converters (VSCs), because of snubberless operation. The high voltage requirement of electrical power transmission and distribution (HVDC) systems is handled by series stacking of IGBTs. As traction systems use parallel connection of devices and HVDC employs series connection, the nature of these applications differs [30]. Consequently, the failure modes of devices in these systems are of opposite nature: *open-circuit failure* for traction and *short-circuit failure* for HVDC. These counter requirements have led to two different packaging concepts: the die-soldered, nonhermetic, wire-bonded module, mounted by bolting to heat sink and single-side cooled; and the dry contact, hermetic, presspack, mounted by pressure stacks and double-side cooled, as adapted from thyristor technology.

1.3 ADVANTAGES AND SHORTCOMINGS OF IGBT

IGBT is created by the functional integration of MOS and bipolar device technologies in monolithic form. It combines the best attributes of the existing families of MOS and bipolar devices (Table 1.2) to achieve optimal device characteristics, approximately fulfilling the criteria of the ideal power switch. Moreover, it has no integral diode like the MOSFET. In an IGBT, absence of the diode provides the user an opportunity to choose an external fast recovery diode suitable for a particular application or to purchase a "co-pak" having the IGBT and the diode in the same package. So, problems associated with the integral diode across the P-base/N-drift region in the power MOSFET are absent in the IGBT.

The major difficulty with the MOSFET is the reverse recovery characteristic of the diode. High carrier lifetime in the N-drift region of as-fabricated MOSFETs makes reverse recovery of diode slow accompanied by a large recovery charge. With increasing voltage ratings, the integral diode exhibits higher reverse recovery charge and reverse recovery time, and thereby high losses. Furthermore, this charge produces a high reverse recovery current, which increases with di/dt. The high current flowing through the transistors in the circuit causes excessive power dissipation and thermal stresses on them. To improve the reverse recovery characteristic, electron irradiation is performed with subsequent annealing of positive oxide charge around 200°C. But still the integral diode in the MOSFET creates problems due to the existence of a bipolar transistor in the structure (Fig.1.3). The voltage drop across the base resistance of this bipolar transistor due to current flowing during reverse recovery forward biases the emitter-base junction of the transistor. The high voltage developed across the transistor often leads to second breakdown. Thus actuation of the bipolar transistor during diode reverse recovery causes serious problems in power MOSFETs.

The IGBT provides high input impedance MOS gating, together with large bipolar current-carrying capability, while designed to support high voltages. A circuit designer views the IGBT as a device with MOS input characteristics and bipolar output characteristics—that is, a voltage-controlled BJT device. This feature simplifies, to a large extent, the driving circuit. This, combined with IGBT ruggedness, eliminates the complexity of protective snubber circuits, allowing simple, lightweight, and economic power electronic systems to be constructed with IGBTs. Over and above, integration of MOS control with bipolar conduction is a way of building intelligence in the chip because "electronic intelligence" is intimately related to the controlling strategy for

Table 1. 2 Features, Pros and Cons of MOSFETs and Bipolars

Serial No.	MOSFETs	Bipolars
	Features:	**Features:**
1	Single-carrier device	Two-carrier device
2	Works by majority carrier drift	Operates by minority-carrier diffusion
3	Voltage driven	Current driven
4	Drain current α channel width	Collector current α emitter length and area
5	Higher breakdown voltage is achieved using lightly doped drain region	Higher breakdown voltage requires lightly doped collector region
6	Current density for given voltage drop is high at low voltages and low at high voltages	Current density for given voltage drop is medium, and severe trade-off exists with switching speed
7	Square-law current–voltage characteristics at low current and linear I–V at high current	Exponential I–V characteristics
8	Negative temperature coefficient of drain current	Positive temperature coefficient of collector current
9	No charge storage	Charge stored in base and collector
	Pros:	**Cons:**
1	High input impedance $Z \sim 10^{9}-10^{11}\ \Omega$	Low input impedance $Z \sim 10^{3}-10^{5}\ \Omega$
2	Minimal drive power. No DC current required at gate	Large drive power. DC current needed at base continuously
3	Simple drive circuit	Complex drive circuit as large positive and negative currents are required
4	More linear operation and less harmonics	More intermodulation and cross-modulation products
5	Devices can be easily paralleled	Devices cannot be easily paralleled
6	No thermal runaway	Prone to thermal runaway
7	Less susceptible to second breakdown	Vulnerable to second breakdown
8	Maximum operating temperature = 200°C	Maximum operating temperature = 150°C
9	Very low switching losses	Medium to high switching losses depending on trade-off with conduction losses
10	High switching speed, which is less temperature-sensitive	Lower switching speed, which is more sensitive to temperature
	Cons:	**Pros:**
1	High ON resistance	Low ON resistance
2	Low transconductance	High transconductance

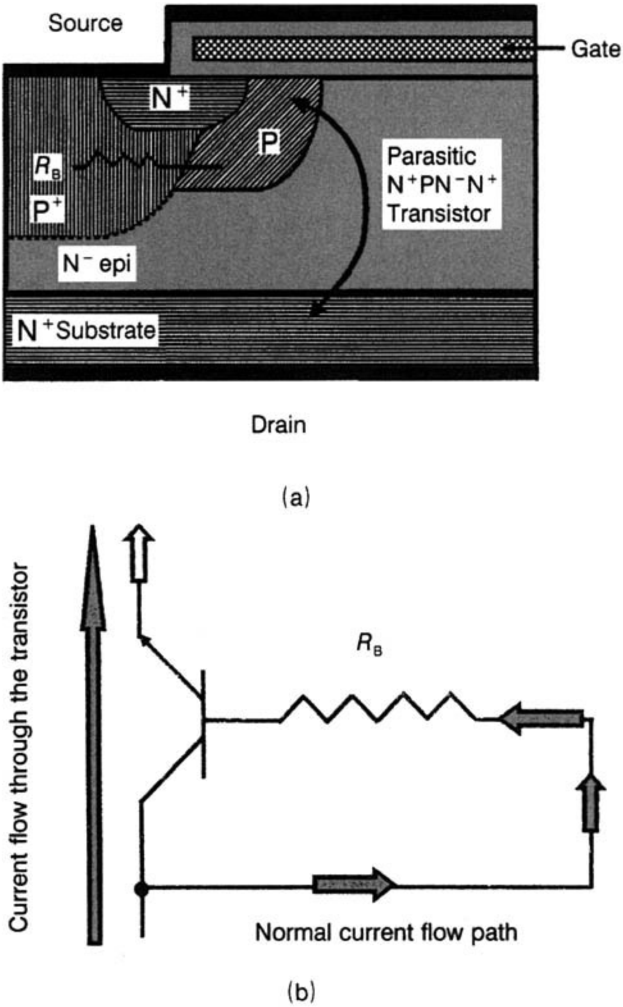

Source
Gate
N⁺
P
R_B
P⁺
Parasitic N⁺PN⁻N⁺ Transistor
N⁻ epi
N⁺Substrate
Drain
(a)
Current flow through the transistor
R_B
Normal current flow path
(b)

Figure 1.3 (a) Parasitic bipolar transistor in power MOSFET structure. (b) Equivalent circuit showing the parasitic bipolar transistor and the resistance R_B of the P base underneath the N^+ source.

switching the power device on and off. Thus it represents a step toward "smart or intelligent power switching." It must not be forgotten here that besides the BJT and the MOSFET, features of P-N junction diodes, P-I-N rectifiers, and the thyristor are also visible in the IGBT.

Let us make a comparative assessment of current-carrying capacities of IGBT with MOSFET and BJT. We note that IGBTs are fabricated in a range of voltage ratings: for example 300-V IGBTs are used for applications based

Table 1.3 Features of IGBT and FET

Property	IGBT	FET
Junctions	Two-junction device	Zero-junction device
Forward voltage	> 0.7 V	> 0 V
Blocking voltage	1200 V	500 V
Forward current	400 A	50 A
Turn-on time	0.9 μsec	90 nsec
Turn-off time	1.5 μsec	150 nsec
Switching frequency	> 150 kHz	> 1 MHz

on rectified 110-V AC line, 600-V IGBTs for rectified 220-V AC line, and 1200-V IGBTs for rectified 440-V AC line. The example of IGBTs with 600-V blocking capability is cited. Its room-temperature current density is 200 A/cm^2, which is 20 times that of the equivalent power MOSFET and 5 times that of the BJT having a current gain of 10 and forward drop of 2 V. At 200°C, the current density of IGBT becomes 60 times that of the MOSFET. To quote another illustrative case, the ON resistance of a 400-V IGBT is 0.1 Ω at 20 A, which is 0.1 that of the ON resistance of an equally rated MOSFET. Not only is the ON resistance of IGBT low, its temperature coefficient (TCR) is much smaller than MOSFET.

Now concerning the disadvantages of the IGBT, the major penalty paid by this hybridization of MOS and bipolar properties, is the slower switching of the IGBT compared to that of the power MOSFET (Table 1.3). Nevertheless, an interesting feature of IGBT is that its turn-off time can be decreased by electron or proton irradiation at the expenditure of an increase in forward drop. This unique capability endows the IGBT with the valuable opportunity of trading off between switching and conduction losses to cater to the power switching requirements of a broad application range. To elaborate, it is noted that for circuits operating at low frequencies with large duty cycles, such as the line-operated phase-control circuits, where conduction losses dominate over switching losses, turn-off times from 5 to 20 μsec are sufficient. For high-frequency circuits with short duty cycles (e.g., AC motor drives working at 1–20 kHz switching frequencies), the necessary turn-off times cover the range 500 nsec to 2 μsec. For still higher frequency circuits (e.g., switching power supplies operating at 20–100 kHz), the required turn-off times lie in the range from 100 to 500 nsec. Suitable IGBTs can be fabricated for all the above applications, compromising between the forward drop and turn-off time specifications.

It must be mentioned here that the IGBT requires a minimum forward voltage of at least 0.7 V before turning on, whereas the MOSFET can conduct with $V_{DS} > 0$ V. Therefore, in low supply voltage applications (e.g., 12-V automotive electronics), IGBT is not a good choice.

1.4 IGBT STRUCTURE AND FABRICATION

Figure 1.4 shows, schematically, the cross section of the basic IGBT structure. This is one of the several structures possible for this device. In practice, the power IGBT essentially comprises a repetitive array of millions of cells arranged, in a topological layout, providing a large aspect ratio: width versus length (W/L). Figure 1.5 represents an *elementary unit cell* of this multicellular power electronic component. The picture in Fig 1.6 shows the unit cell appearance in three dimensions. The different device regions of the structure must be clearly recognized. The N^+ layer at the top is the emitter. The P^+ layer at the bottom constitutes the collector. It is also called the *hole-injecting layer*. There are two base regions known as the P *base* and N *base*. Essentially, the IGBT comprises a four-layer N-P-N-P thyristor structure.

From Figs. 1.5 and 1.6, it is evident that the silicon cross section of the IGBT is virtually similar to the vertical DMOSFET (VDMOSFET) except for the P^+ substrate in the IGBT. Also, both devices have a polysilicon gate structure and P wells with N^+ source regions. The thickness and resistivity of the N-type material between P wells controls the voltage rating of both the devices. Therefore, IGBT is fabricated by an N-channel polysilicon-gate self-aligned vertical DMOSFET (VDMOSFET) process. Because the IGBT

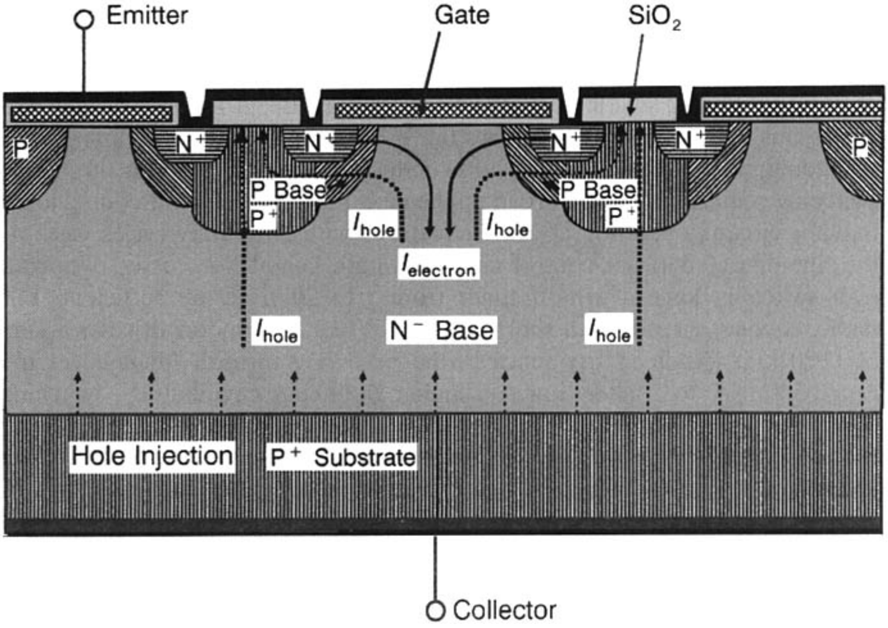


Figure 1.4 Schematic cross-sectional view of the IGBT.

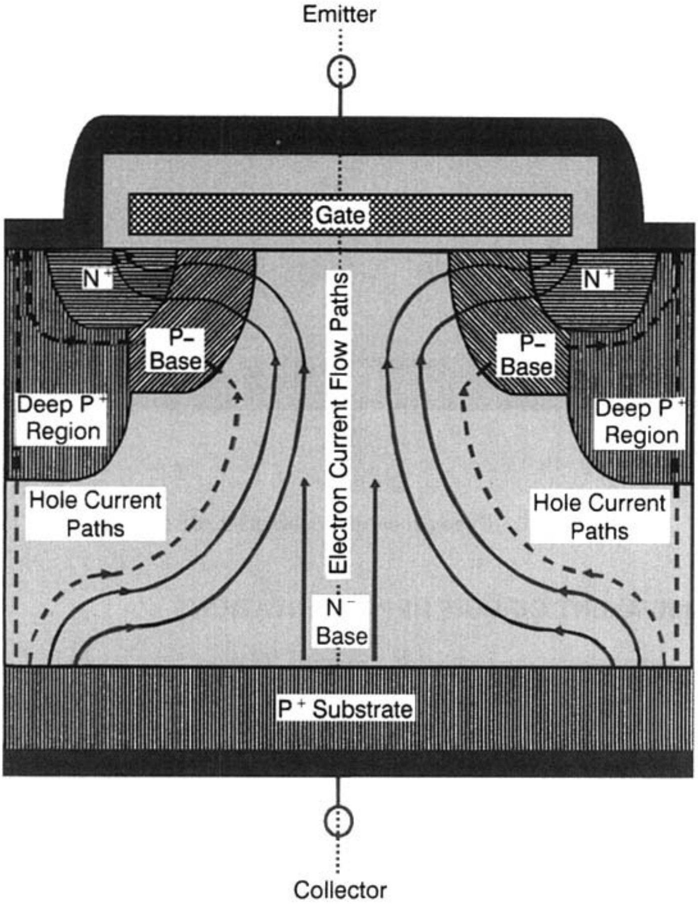

Emitter
Gate
N+
N+
P–
Base
P–
Base
Deep P+
Region
Deep P+
Region
Electron Current Flow Paths
Hole Current
Paths
Hole Current
Paths
N−
Base
P+ Substrate
Collector

Figure 1.5 IGBT unit cell. Half unit cell obtained by cutting across the dotted line represents the basic building block for analysis.

is essentially a VDMOSFET with an additional PN junction in the drain region, the process sequence of IGBT fabrication is the same as for the power VDMOSFET except that the starting N^+ substrate in VDMOSFET is replaced by P^+ substrate in IGBT. Like the SCR, an IGBT is fabricated with Si material due to its good thermal conductivity and high breakdown voltage. N^- epitaxial layer is grown over the starting P^+ substrate followed by ion implantation, thermal diffusion, oxidation, chemical vapor deposition, and photolithographic steps as in poly-Si gate MOSFET processing. The details of fabrication technology will be described in Chapter 8.

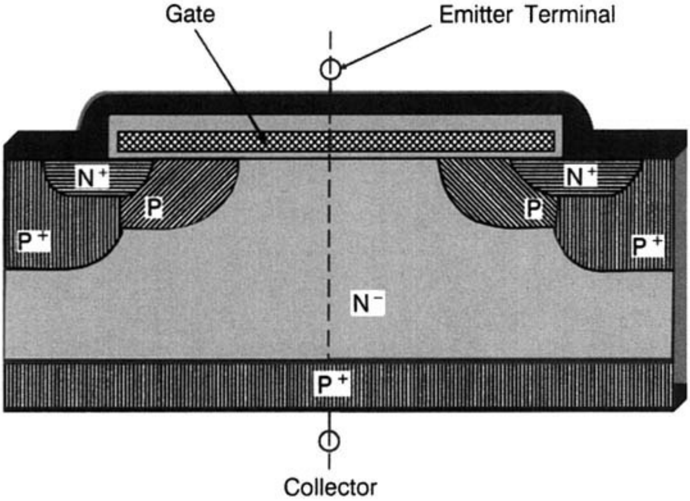


Figure 1.6 Three-dimensional view of the IGBT unit cell.

1.5 EQUIVALENT CIRCUIT REPRESENTATIONS

A clear perception of IGBT operation can be obtained by examining its equivalent circuit. From a close inspection of the IGBT cross-sectional diagram in Fig. 1.4, its equivalent circuit model is drawn in Fig. 1.7a. A pair of NPN and PNP transistors represents the thyristor. The collector of the NPN transistor is connected to the base of the PNP transistor, and likewise the collector of PNP transistor supplies the base current for the NPN transistor through the JFET. The NPN and PNP transistors thus constitute a regenerative feedback loop. It can also be seen from Fig. 1.7a, that sintered aluminum metallization as well as deep P^+ diffusion in the center of P base short-circuit the emitter and base of the NPN transistor. This is represented by the shorting resistance R_s connecting the emitter and base of NPN transistor in Fig. 1.7a. Emitter-base shorting is essential to ensure that the sum-total gain of NPN and PNP transistors ($\alpha_{\mathrm{NPN}} + \alpha_{\mathrm{PNP}}$) does not exceed unity so that the thyristor does not latch up. Latching is avoided because it results in loss of gate control over output current. The latch-free IGBT, thus obtained, behaves as a single bipolar transistor in the form of its PNP component. Furthermore, the MOSFET channel is formed in the P base below the gate oxide. This channel joins the N^+ emitter and N^- collector of the NPN transistor. So, this transistor is shunted by a MOSFET in the equivalent circuit. The JFET in the circuit represents the constriction of current between any two neighboring IGBT cells.

For proper IGBT functioning, it is obvious that the operation of the NPN transistor is deliberately avoided. Hence, the NPN transistor can be neglected. Then Fig. 1.7a is replaced by the simplified equivalent circuit

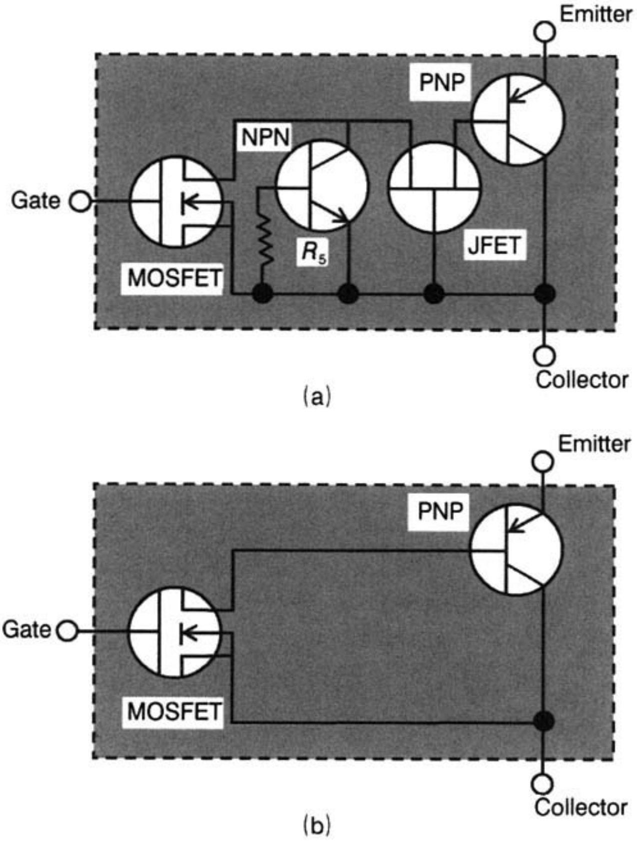


Figure 1.7 Electrical equivalent circuit models of IGBT. (a) Equivalent circuit of IGBT. (b) Simplified equivalent circuit of IGBT.

illustrated in Fig.1.7b. This circuit contains only two elements—for example the MOSFET and the PNP transistor. Therefore, the IGBT is viewed as an N-channel enhancement mode MOSFET driving a PNP bipolar transistor in a *pseudo-Darlington configuration.* Being the final stage of a pseudo-Darlington transistor, the PNP transistor is never in heavy saturation under normal operation so that its voltage drop is always higher than that of a saturated PNP transistor. The condition under which the PNP transistor will saturate is during latchup of the IGBT, which is an undesirable condition from which the device must be kept away. But the P^+ layer in an IGBT covers the full chip area. Hence its injection efficiency and conduction drop are superior to that of a BJT of equal size. The fact that PNP transistor does not operate under saturation gives the IGBT advantage during turn-off because it is easier to switch off an unsaturated transistor than a saturated one. Furthermore, because the base of PNP is not externally accessible, the turn-off time cannot be reduced by a drive circuit. Lifetime killing techniques and N-buffer layer are used for this purpose. But they decrease the current gain of PNP transistor and increase the forward drop of IGBT. The gain of PNP transistor therefore controls the conduction and switching losses. It also determines the latchup immunity of the built-in thyristor in IGBT as follows. The N-buffer layer and the wide epi N^- base decrease the gain of the PNP transistor while the gain of the NPN transistor is reduced by decreasing its base resistance, thereby inhibiting dynamic latching during turn-off. During that period, a large hole current density flows through its P base, raising its gain to high values.

From these considerations, it is evident that the design of an IGBT involves the optimization of both the MOSFET and the bipolar parts. Naturally, there are two approaches to reduce the conduction drop of the IGBT: either (i) to decrease the ON resistance of the MOSFET by increasing the chip size and cell density, resulting in a "conductivity-modulated MOSFET" or (ii) to increase the current gain of the PNP transistor (with due attention to the latching and blocking voltage capabilities), producing a "MOSFET-driven bipolar transistor." Both approaches are applied in practice.

Considering the IGBT as a PIN rectifier whose output current flows through a MOSFET channel provides further simplification, but this is not an accurate representation of the equivalent circuit. Therefore, it is generally not used as IGBT equivalent circuit.

1.6 PRINCIPLE OF OPERATION AND CHARGE-CONTROL PHENOMENA

With reference to Figs. 1.4, 1.5, and 1.6, and with zero gate bias applied, the IGBT structure is equivalent to a PNPN break-over diode having a emitter short. The IGBT remains off when the collector is positively biased with respect to the emitter ($V_{CE} > 0$). This is because the junction between the P-base and N^--type epitaxial layer is reverse-biased. Similarly, the IGBT remains in the off-state, when $V_{CE} < 0$, due to reverse bias across the junction between P^+ substrate and N^- epitaxial layer. So, the collector-emitter current I_{CE} is minimal until breakdown in both polarities. For forward V_{CE} the breakdown is caused by avalanche of the N^-P junction, while for reverse V_{CE} it is initiated by the same process at the N^-P^+ junction.

Applying the $V_{CE} > 0$ condition, the IGBT is turned on by applying a positive gate-emitter voltage (V_{GE}) of sufficient magnitude to induce an N channel in the underlying P region, thereby connecting the N^+ emitter with N^- base, forward biasing the base-emitter junction of the PNP transistor and making it conduct. The current flow across the junction between the N^--type epitaxial layer and the P-type substrate results in injection of minority-carrier holes into the epitaxial layer. Consequently, conductivity modulation of this layer takes place, reducing its resistance and bringing about a large flow of collector–emitter current I_{CE}. To turn-off the IGBT, the gate–emitter voltage V_{GE} is made zero so that the channel in the P region is removed. Depending on the value of V_{CE}, three different regions of operation are observed in the IGBT. In the *first regime*, at a small value of $V_{CE} \sim 0.7$ V, the IGBT is a VDMOSFET in parallel with a PNP transistor. Current transport takes place by recombination of excess electrons and holes in N^--region. The *second regime* commences from $V_{CE} > 0.7$ V where the characteristics portray MOSFET behavior. At high V_{CE} values, the excess holes injected from the emitter of the PNP transistor are not absorbed by recombination in the N^- base and spill over to the P base contributing to PNP bipolar current. The MOSFET current I_{MOS} is the base current of the bipolar transistor, and the collector–emitter current I_{CE} of IGBT duplicates the general shape of MOSFET characteristics except that it is the amplified version of I_{MOS}. In the *third regime*, when the current exceeds a critical level, the device latches up like the ON state of a thyristor. Consequently, gate control is lost.

Device operation will be dealt with at length in Chapter 2. It may be remarked here that the MOS part of the IGBT controls its turn-on while the bipolar part determines the steady-state and turn-off behavior. Like the BJT and the MOSFET devices, the IGBT is also a charge-control device. In a PNP BJT operating in the active mode, the electron injection from the base terminal produces a negative charge in the N^- base. This is balanced by an injection of holes from the P-emitter for maintaining charge neutrality. So the hole concentration in the base region is enhanced, allowing the holes to move from emitter to collector. Similarly, in an N-channel MOSFET, the conduction begins when its input capacitance has received adequate positive charge to raise the gate–source voltage to the level required to achieve inversion for channel formation. The main difference between bipolar and MOSFET conduction is that in the BJT a continuous supply of base current in the form of electrons is imperative to replenish the electrons lost by recombination. But in the MOSFET the gate dielectric (oxide) separating the electrons (in the channel) and positive charge on the oxide prevents their recombination so that current flow is necessary only for establishing charge density or during its withdrawal. Also, this charge requirement is very small. It follows that the IGBT that merges both bipolar and MOSFET properties is also a charge-control device.

1.7 CIRCUIT MODELING

Device manufacturers and circuit designers require IGBT models to understand device internal mechanisms, optimize structures, and predict circuit behavior [31–53]. Various circuit simulation packages—notably, Saber and members of the Simulation Program with Integrated Circuit Emphasis (SPICE, PSPICE, HSPICE, IG-SPICE, etc.)—are available for IGBT modeling. Models are classified as mathematical (analytical models based on semiconductor physics), semimathematical (combining device physics with existing models in the simulator), behavioral or empirical (simulating IGBT characteristics without consideration of physical mechanisms), and seminumerical (using finite element methods to model the wide base and using analytical methods for other device parts). They are also categorized as microscopic (based on physical structure and equations) and macroscopic or composite (utilizing the available device models).

The macroscopic IGBT model, following the semimathematical approach, has been formulated in SPICE based on the existing models of MOSFET and BJT devices. This model is useful for circuit simulation, and it is less complicated and time-consuming than the microscopic model. Figure 1.8a presents the resistive or DC model of IGBT in SPICE. Three additional voltage- and current-controlled generators can be seen between the MOSFET and BJT sections. These are included because the output characteristics of the IGBT are not exactly the same as those of the MOSFET, calling for modifications. Furthermore, since SPICE requires that the inputs of controlling variables must be of only one source type, either all voltage or all current, the controlling variables are converted into the same type.

Diode D and the two-dimensional voltage-controlled voltage source (VCVS) E_D represent the typical value of output voltage (0.7–1.0 V). The dimension of a source is understood from the definition of *non-linear-dependent sources* in SPICE which are characterized by any of the four equations $i=f(v)$, $v=f(v)$, $i=f(i)$ and $v=f(i)$, where the functions are polynomials and the arguments can be multidimensional. The diode prevents the collector current of the BJT from flowing on applying a reverse voltage across the output of IGBT. H_D is a current-controlled voltage source (CCVS), which transforms the drain current of the MOSFET into the driving voltage V_D across the resistor R_D. G_B is a two-dimensional nonlinear voltage-controlled current source, which drives the output transistor.

Figure1.8b shows the dynamic SPICE model of IGBT, which is a combination of the DC model and the nonlinear input capacitance C_{ix}, modeled by a four-segment piecewise linear function. The reverse capacitance is very small while the output capacitance C_o is taken from the data sheets. The current-controlled input capacitance is synthesized with the nonlinear input resistive circuit to obtain the complete SPICE macromodel of the IGBT.

The accuracy of this model is lower than that of the mathematical IGBT models because the wide base in IGBT differs from the existing discrete power BJT models. IGBT switching behavior is primarily controlled by the distributed charge in this wide base. The behavior of the distributed charge is governed by the ambipolar transport equation, a second-order partial differential equation.

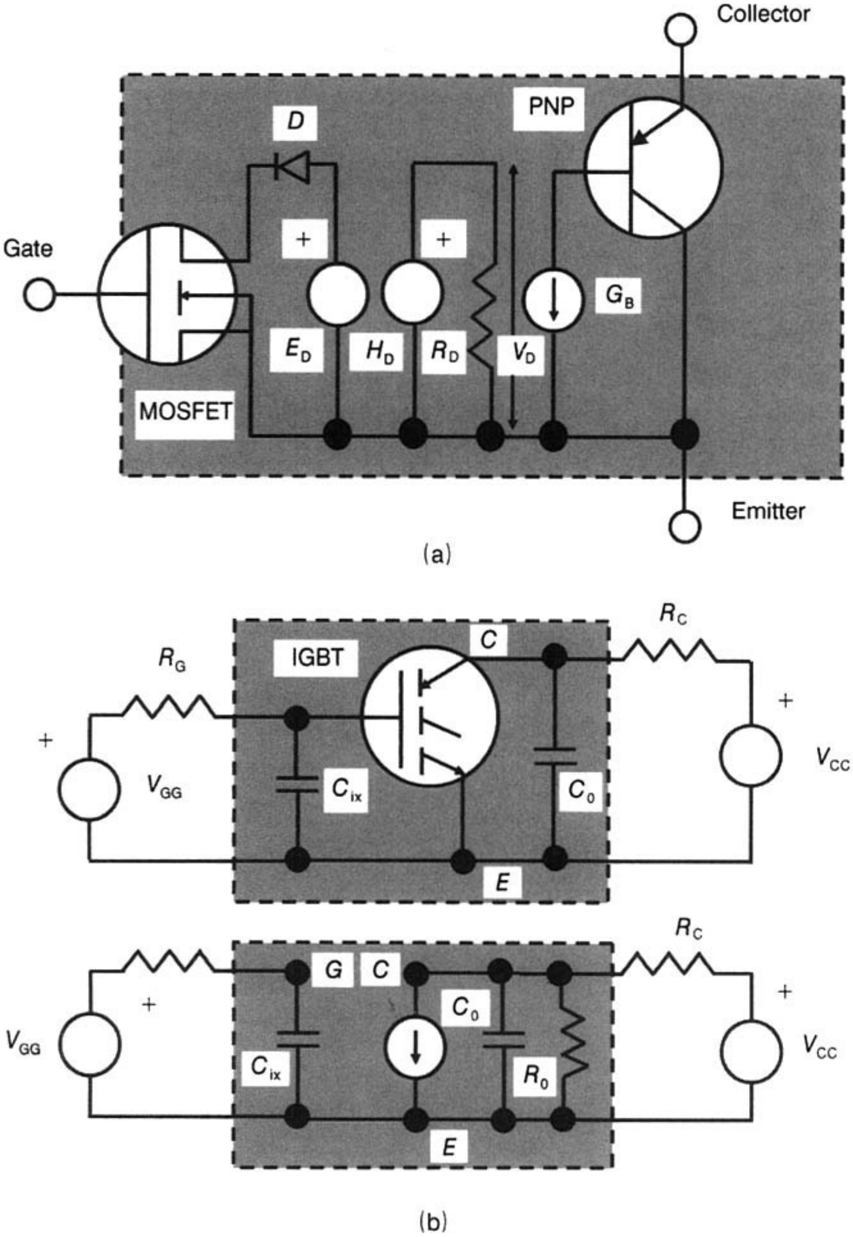

Collector
PNP
D
Gate
+
+
G_B
E_D
H_D
R_D
V_D
MOSFET
Emitter
(a)
R_C
R_G
IGBT
C
+
+
V_{CC}
V_{GG}
C_{ix}
C_0
E
R_C
G
C
C_0
+
+
V_{GG}
C_{ix}
R_0
V_{CC}
E
(b)

Figure 1.8 SPICE models of IGBT. (a) DC model of IGBT. (b) Switching circuit and dynamic model of the IGBT.

Example 1.1 Manufacturing process of IGBTs produces two different types as follows:

IGBT Type	Forward Voltage Drop at 80 A (in volts)	Turn-Off Loss per Switching (in mJ)
Slow IGBT	1.2	10
Fast IGBT	2.1	5

Considering a duty ratio of 50% for a DC drive application, at what frequency will the two IGBTs give the same total power loss? Compare the power losses of the two IGBTs at 5 kHz and 10 kHz.

Neglecting the turn-on, driving, and nonconducting state losses, we can write

$$\text{Power loss } P_{\text{loss}} = \text{Forward voltage } V_F \times \text{Forward curren } I_F \times \text{Duty Ratio} + \text{Switching frequency } (f) \times \text{Energy lost during turn-off} \quad \text{(E1.1.1)}$$

$$\text{For the slow IGBT} \quad P_{\text{loss}} = 1.2 \times 80 \times 0.5 + f \times 10^3 \times 10 \times 10^{-3} = 48 + 10f \quad \text{(E1.1.2)}$$

$$\text{For the fast IGBT,} \quad P_{\text{loss}} = 2.1 \times 80 \times 0.5 + f \times 10^3 \times 5 \times 10^{-3} = 84 + 5f \quad \text{(E1.1.3)}$$

The power losses of the slow and fast IGBTs will be equal at a frequency f(kHz) obtained from the equality

$$48 + 10f = 84 + 5f \quad \text{(E1.1.4)}$$

from which $f = 7.2$ kHz.

At f = 5 kHz,

For the slow IGBT, $$P_{\text{loss}} = 1.2 \times 80 \times 0.5 + 5 \times 10^3 \times 10 \times 10^{-3} = 98 \text{ W} \quad \text{(E1.1.5)}$$

For the fast IGBT, $$P_{\text{loss}} = 2.1 \times 80 \times 0.5 + 5 \times 10^3 \times 5 \times 10^{-3} = 109 \text{ W} \quad \text{(E1.1.6)}$$

At f = 10 kHz,

For the slow IGBT, $$P_{\text{loss}} = 1.2 \times 80 \times 0.5 + 10 \times 10^3 \times 10 \times 10^{-3} = 148 \text{ W} \quad \text{(E1.1.7)}$$

For the fast IGBT, $$P_{\text{loss}} = 2.1 \times 80 \times 0.5 + 10 \times 10^3 \times 5 \times 10^{-3} = 134 \text{ W} \quad \text{(E1.1.8)}$$

Thus we conclude that at 5 kHz, the slow IGBT is superior, giving a lower power loss (98 W) than the fast IGBT (109W). At 10 kHz, the slow IGBT becomes inferior because it gives a higher power loss (148 W) than the fast IGBT (134 W).

Example 1.2 Assuming a 50% duty cycle, calculate the power dissipation of 1000-V IGBT, BJT, and MOSFET chips of equal area, each carrying a current of 50 A, given the following set of parameters for the devices:

Serial No.	Device	Forward Voltage Drop at 50 A (volts)	Turn-Off Time (μsec)
1	IGBT	2	1
2	BJT	15	1
3	MOSFET	40	0.1

Perform these calculations at two frequencies, 25 kHz and 100 kHz.

Neglecting turn-on, driving, and OFF-state losses, the power dissipation P_D is the sum of power loss incurred during steady-state forward conduction (P_{ss}) and power loss during switching from ON state to OFF state (P_{sw}), that is

$$P_D = P_{ss} + P_{sw} \tag{E1.2.1}$$

Let T be the periodic time of the gate pulse, τ_1 the time fraction of the time period T for which the device remains on, τ_2 the turn-off time, V_F the forward voltage drop at current I_F, and V_B the blocking voltage. Then the power dissipation is expressed as

$$P_D = V_F I_F \tau_1 + (1/2) V_B I_F (\tau_2/T) \tag{E1.2.2}$$

The first term (P_{ss}) in this expression is independent of frequency and depends only on the duty cycle so that we have the following:

For IGBT,

$$P_{ss} = 2 \times 50 \times 0.5 = 50 \text{ W}$$

For BJT,

$$P_{ss} = 15 \times 50 \times 0.5 = 375 \text{ W}$$

For MOSFET,

$$P_{ss} = 40 \times 50 \times 0.5 = 1000 \text{ W}$$

Total power dissipation at 25 kHz is as follows:

For IGBT,

$$P_D = 50 + (1/2) \times 1000 \times 50 \times 1 \times 10^{-6} / \{1/(25 \times 10^3)\} = 675 \text{ W}$$

For BJT,

$$P_D = 375 + (1/2) \times 1000 \times 50 \times 1 \times 10^{-6} / \{1/(25 \times 10^3)\} = 1000 \text{ W}$$

For MOSFET,

$$P_D = 1000 + (1/2) \times 1000 \times 50 \times 0.1 \times 10^{-6} / \{1/(25 \times 10^3)\} = 1062.5 \text{ W}$$

Total power dissipation at 100 kHz is as follows:

For IGBT,

$$P_{\mathrm{D}} = 50 + (1/2) \times 1000 \times 50 \times 1 \times 10^{-6}/\{1/(100 \times 10^3)\} = 2550\ \mathrm{W}$$

For BJT,

$$P_{\mathrm{D}} = 375 + (1/2) \times 1000 \times 50 \times 1 \times 10^{-6}/\{1/(100 \times 10^3)\} = 2875\ \mathrm{W}$$

For MOSFET,

$$P_{\mathrm{D}} = 1000 + (1/2) \times 1000 \times 50 \times 0.1 \times 10^{-6}/\{1/(100 \times 10^3)\} = 1250\ \mathrm{W}$$

From the above, it is irresistible to conclude that at 5 kHz the IGBT is the least lossy of the three devices. At 10 kHz the IGBT still remains better than BJT but both IGBT and BJT become inferior to MOSFET due to increased power losses.

1.8 PACKAGING OPTIONS FOR IGBTs

Desirable characteristics of a package include good electrical and thermal performance, long life, high reliability, and low cost. Additionally, for a module, electrical isolation of the base plate from the semiconductor die is necessary in order that both the halves of a phase leg are enclosed in one package. This is also required for convenience so that modules switching, different phases, can be mounted on the same heat sink, and also from safety considerations for grounding the heat sink. IGBTs are generally packaged in three types of commercially available packages (Fig. 1.9): (i) *Discrete packages* such as TO-220, TO-247, TO-264, and SOT-227B. These packages contain a single device and are used for low-power applications. (ii) *Power module package*: These package multiple dice and are available in several configurations such as half-bridge, full-bridge, and three-phase bridge. (iii) *Press pack*: This type has been recently introduced for modules (see Chapter 9, Section 9.7).

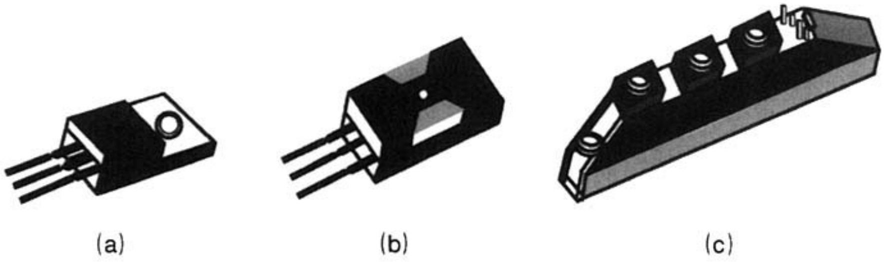


Figure 1.9 Common IGBT packages. (a) TO-220 AB. (b) TO-247 AD. (c) TO-240 AA.

1.9 HANDLING PRECAUTIONS OF IGBTs

(i) Before assembly into a circuit, the IGBTs should be stored with their leads shorted together by metallic springs or kept in a conducting material. (ii) The tip of the soldering iron used must be grounded. (iii) When touched by hand, a metallic wristband must ground the hand. (iv) A zener diode may be connected between gate and emitter, for gate protection from electrostatic charge buildup. (v) The specified gate–emitter voltage should not be exceeded. Circuits in which the gate terminal is kept floating must be avoided.

1.10 IGBT GATE DRIVING CIRCUITS

The IGBT is a three-terminal device represented by the circuit symbols given in Fig. 1.10. The three terminals are designated as Emitter (E), Gate (G), and Collector (C). Many authors have adopted the terminology of anode/cathode/gate from SCR, while others use drain/source/gate from MOSFET. Some prefer collector/emitter/gate from the BJT technology because the device is a bipolar transistor. The terminal E is actually the collector of PNP transistor, and terminal C is actually the emitter of the PNP transistor. In BJT terminology, the terminal E will be the emitter of IGBT but collector of PNP transistor. Similarly, the terminal C will be collector of IGBT but emitter of PNP transistor. To avoid any confusion resulting from whether we are referring to the emitter/collector of the IGBT or that of the PNP transistor, the anode/cathode/gate terminology seems to be more logical from this viewpoint. However, the collector/emitter/gate terminology from BJT will be adopted in this book to constantly remind the reader that the IGBT is a bipolar transistor, not a thyristor. This is also followed in most data books. The symbol in Fig. 1.10a is most realistic because it represents

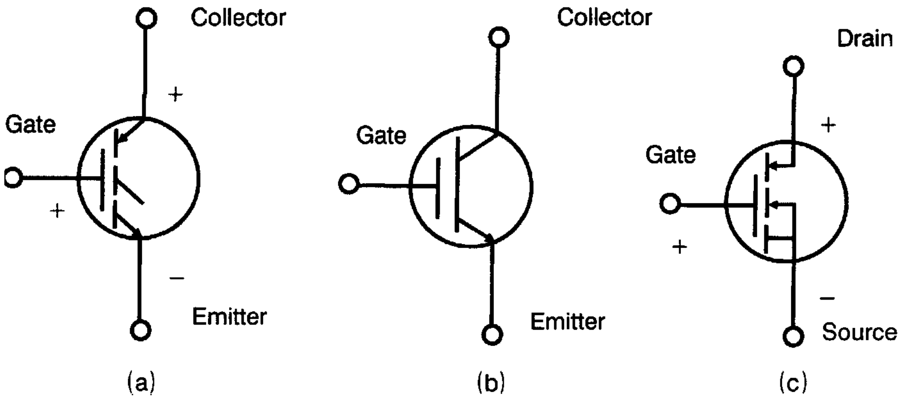


Figure 1.10 Circuit diagram symbols of IGBT.

both the MOSFET gate and the collector and emitter of the PNP transistor in the IGBT. The base contact of the PNP transistor is left unconnected, signifying that there is no external base lead. Externally, the IGBT has collector, emitter, and gate terminals. The symbol in Fig. 1.10b may be misunderstood because the PNP transistor, and not the NPN transistor, in the IGBT plays the major role. This representation is only symbolic and is not related to the actual IGBT structure. The symbol in Fig. 1.10c emphasizes the presence of MOSFET in IGBT. It is identical to the symbol of N-channel MOSFET. The difference between the IGBT symbol and the MOSFET symbol is the arrow on the drain contact. This arrow represents the emitter of the PNP transistor in the IGBT. For a P-channel IGBT, the direction of the arrow is reversed.

The gate power requirements of IGBT are similar to those of power MOSFET, and the gate drive circuits of power MOSFETs are simplest of all the power electronic semiconductor switches. A gate-to-emitter voltage v_{GE} of 15 V is applied to turn on the IGBT. Then v_{GE} is reduced below about 3–4 V to turn off the device. Although the IGBT turn-on and turn-off are controlled by *gate voltage*, the speed of switching is determined by *gate current* because the time for development of the required charge for inversion, depends on the magnitude of gate current.

There are two types of gate driving circuits, namely, voltage drive and current drive. Figure 1.11a shows a voltage drive circuit for IGBT. It consists of a constant voltage supply $V_G = 15$ V, a switch, and two resistors R_1 and R_2. The switch comprises a fast-switching transistor or a complete IC chip including the interfacing between it and the control signal from a microprocessor-based controller. As soon as the switch is closed, a gate current i_G flows until the input gate capacitance C_{iee} of the IGBT is charged to a steady-state voltage $v_{GE} = V_{GE}$ when $i_G = 0$. The time taken for this charging operation measures the time for initiation of turn-on and is decided by the values of C_{iee}, R_1, R_2 and R_G (the internal resistance of the gate supply V_G). A smaller value of the resistance $(R_G + R_1)$ results in a shorter charging time constant. Since R_G is generally not variable, the resistor R_1 is decreased to provide faster charging and hence rapid turn-on of IGBT. Connection of an additional capacitor C_1 across R_1 can speed up the turn-on process still more. Upon opening the switch, the gate supply V_G is disconnected and the gate capacitance C_{iee} discharges into the resistor R_2. When the gate voltage has fallen to a value $v_{GE} < V_{Th}$, the IGBT turns off. It is readily apparent that the turn-off process can be hastened by decreasing the resistance value R_2 to enable quicker discharging.

The circuit diagram of a gate driving circuit using a constant current gate supply I_G is shown in Fig. 1.11b. As soon as the switch is opened, a constant gate current $i_G = I_G$ flows into the gate terminal of IGBT. This continues until the voltage across the IGBT reaches the breakdown voltage of the zener diode, $V_Z = 15$ V. Then the gate current, i_G, becomes zero and the zener diode conducts until the closure of the switch. Upon closing the switch, the

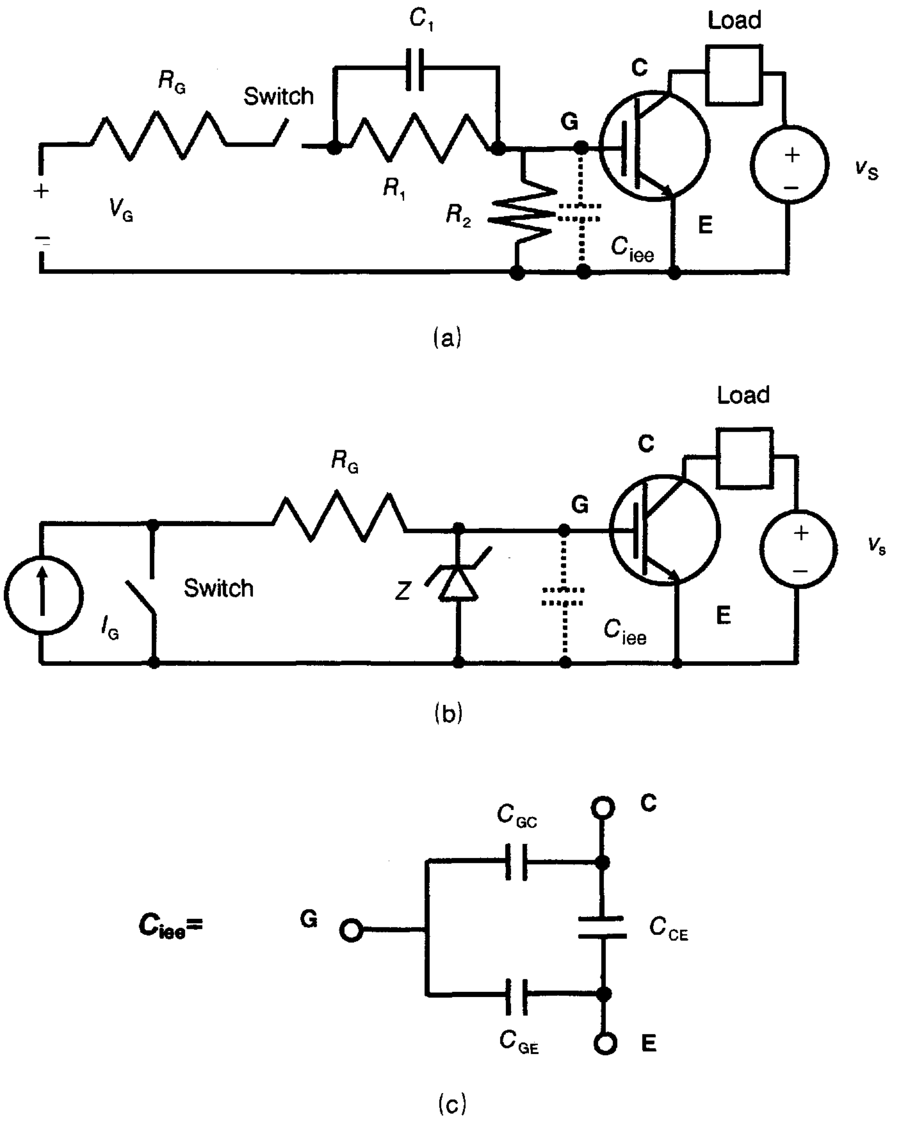

C1
Load
RG
Switch
C
G
+
VG
R1
R2
Ciee
E
VS
(a)
Load
C
RG
G
Switch
Z
IG
Ciee
E
Vs
(b)
CGC
C
Ciee=
G
CCE
CGE
E
(c)

Figure 1.11 Gate drive circuits of IGBT and components of the input capacitance. (a) Voltage gate drive circuit of IGBT. (b) Current gate drive circuit of IGBT. (c) Components of C_{iee}.

IGBT gate discharges through the resistor R_G so that the closed switch acts as a path for the flow of both the discharge current and the supply current I_G. When the gate voltage decreases and becomes less than the threshold voltage V_{Th}, the IGBT turns off.

1.11 IGBT PROTECTION

All semiconductor switches need protection against the damaging overvoltages, overcurrents and transient surges. The IGBT is no exception. IGBT protection circuits are relatively less complex than for BJTs and thyristors, primarily because of the fact that it is a voltage-controlled device [54, 55]. Upon quick detection of these detrimental situations, the IGBT can respond by removing or reducing the gate drive voltage.

Overvoltages. In the forward conduction state, overvoltages do not have any harmful effect because the IGBT acts as a short circuit. But during the blocking state, any applied voltage in excess of the breakdown voltage between collector and emitter terminals causes avalanche breakdown. Consequently, uncontrolled high currents flow, producing appreciable power dissipation and device heating, which may cause permanent damage. The precautionary measures adopted against overvoltages include limiting the supply voltage to less than 80% of the rated breakdown voltage of IGBT and connecting a nonlinear *voltage arrester* in parallel with the IGBT. The voltage arrester is a metal-oxide varistor whose resistance value falls as voltage rises, thus providing a short-circuit path at high voltages. If the reverse blocking voltage of the IGBT is small, a reverse-connected diode placed across the IGBT is helpful. However, any stray inductance L_s in series with the diode must be minimized because it will apply a reverse-bias voltage $= L_s\,(di_{\text{diode}}/dt)$ across the IGBT when the load current is diverted through the diode due to a switching action elsewhere in the circuit. A diode connected in series with the IGBT is sometimes used; but its drawback is that it adds a voltage drop during forward conduction, thereby decreasing the efficiency of power modulation. Several IGBTs may be connected in series to withstand a voltage much higher than that of a single device.

The IGBT gate is sensitive to static charge buildup. Proper grounding is necessary to prevent this deleterious charge accumulation during handling. Normally, the gate voltage should not exceed ± 20 V otherwise the gate oxide will be punctured. A zener diode of breakdown voltage < 20 V, with adequate current rating, connected across the gate-to-emitter terminals, is able to limit the gate voltage to a safe level.

Overcurrents. Like any power device, the rated current of an IGBT is the continuous current for which the junction temperature will not exceed 150°C if the device case temperature is 25°C. So an overcurrent is the current value at which the junction temperature becomes more than 150°C. Such a current value is obviously not permissible. However, pulsed currents much higher than continuous currents are allowed, because the thermal limitation is governed by average power dissipation, and the dissipation in pulse mode is relatively less. Protection from overcurrents is implemented through *thermal fuses* or *gate control circuits* to remove or reduce the gate voltage and turn-off the IGBT when the situation warrants so. When an overcurrent is sensed (e.g., by means of voltage drop across a resistor), the gate voltage is decreased by switching in a zener diode across the gate terminals within a short time interval of ~ 1 μsec. If the fault is cleared away in this time span, normal gate voltage can be reapplied. However, quick interruption of a current produces an Ldi/dt voltage spike, so protection against the spike must not be overlooked. If the load current is far above the current capability of a single IGBT, several IGBTs may be connected in parallel to share the current.

It must be mentioned that the IGBTs are not plagued by second breakdown as are BJTs and have a large safe operating area. The current flow in an IGBT is self-limiting to some extent because if a fault condition abnormally increases the current, the device may leave the ON state and move to the active region where the current is unvarying and independent of the collector-emitter voltage.

Transients. The initial inrush currents at IGBT turn-on arising from motor starting, snubber capacitor discharge, or reverse recovery of a diode across a load must be overcome. For this purpose, a ramp waveform or a two-step waveform in which the first step is from 5 to 7 V and the second step up to 15 V replaces the conventional stepped voltage waveform. By this strategy, not only is the rate of rise of current limited, the peak value of current is also smaller.

IGBTs can turn on by dv_{CE}/dt, either because the parasitic thyristor begins to conduct or the voltage v_{GE} increases above threshold by capacitor division; the capacitors C_{GE} and C_{GC} act as a voltage divider. The latter is avoided by placing a low resistance between gate and emitter terminals. For protection against dv_{CE}/dt effect, a negative gate bias supply $-V_{\mathrm{G}}$ is used during turn-off and in the OFF state.

1.12 SUMMARY AND TRENDS

The IGBT is a fully controllable switch that has carved a niche for itself in medium-power and medium-frequency chopping, conversion and inversion applications, where BJTs and MOSFETs have been traditionally exploited. It is well-suited for general-purpose switching to drive lamps, heaters, solenoids, and motor drives, especially for pulse width modulation. Although it possesses both forward and reverse blocking capabilities, device designers often sacrifice the reverse blocking in favor of forward voltage drop or switching speed. An auxiliary diode is connected antiparallel with the IGBT, either *internally* within the device packaging or *externally* to support the reverse voltage. The IGBT is free from second breakdown problems. In opposition to the IGBT, the MOS-SCR [56] and MCT [57–69] have so far not been able to make the desired impact on the power electronics scenario, and presently their importance in the market is very low. No extensive range of MCTs has so far been released. Its frequency response resembles the IGBT. But its lower conduction losses and higher operational current density bestow an enormous potential in foreseeable future for replacing SCRs and GTOs.

After acquiring a perspective of the different types of power semiconductors and MOS-bipolar hybrid devices in this chapter, the ensuing chapter will focus on the various IGBT structures vis-à-vis their performance and will delve into the physical principles of IGBT operation.

REVIEW EXERCISES

1.1 List the essential prerequisites of an ideal power semiconductor switch.

1.2 How are the following losses in a power semiconductor switch minimized: (a) ON-state losses, (b) OFF-state losses, (c) switching losses, and (d) driving losses? Give reasons for your answers.

1.3 Name the currently available highest current density device. Does it provide gate turn-off feature?

1.4 What is the relative performance of BJT and GTO thyristor as power devices with a controlling base/gate.

1.5 Why are power MOSFETs unsuitable for high-voltage applications? For what types of applications are they most appropriate? Discuss with examples.

1.6 Write two advantages of MOSFETs over bipolar transistors. Give two advantages of bipolar transistors over MOSFETs.

1.7 Draw and explain the circuit diagrams of the following composite device configurations: (a) Darlington, (b) cascade, and (c) cascode.

1.8 What does the acronym IGBT stand for? Mention two other abbreviated representations for this device along with their full forms.

1.9 Name two other members of the IGBT family of MOS-bipolar combination devices.

1.10 Distinguish between the functional capabilities of the MOS-gated SCR and MCT.

1.11 State whether the PNP transistor in the IGBT equivalent circuit is a low-gain or high-gain transistor? Under what injection conditions, low-level or high-level, does this transistor work for the practical operating current density range of IGBT? Can the transport of electrons and holes in this transistor be treated independently? What type of transport equations, unipolar or ambipolar, are used for this analysis?

1.12 What is the structural difference between vertical IGBT and VDMOSFET? Explain how this structural difference is translated into the difference of output characteristics of these devices?

1.13 (a) What are the advantages of MOS gate in IGBT? (b) How does bipolar conduction decrease the forward voltage drop of an IGBT? (c) Explain "Conductivity Modulation of N^- base of an IGBT" and its effect. (d) How does *minority-carrier storage* in N^- base of an IGBT lengthen its turn-off time? (e) Comment on the statement, "IGBTs can be fabricated for a range of forward voltages and turn-off times." How does this aid in controlling the switching losses?

1.14 (a) Draw the schematic of IGBT cross section and label its parts/terminals. Draw also the cross-sectional diagram of a power DMOSFET. How does it differ from that of the IGBT? (b) Explain the operating principle of IGBT. Is it justified to view it as a *charge-control device*?

1.15 Draw and explain three circuit diagram symbols of the IGBT. Which symbol would you like to use and recommend?

1.16 Give one application example each of IGBTs having the turn-off times in the following ranges: (a) 5–20 μsec, (b) 0.5–2 μsec, and (c) 0.1–0.5 μsec.

1.17 The ON-state voltage drop of an IGBT A at 100 A is 1.5 V, while that of IGBT B at the same current is 2.5 V. Their turn-off losses per switching are, respectively, 8 mJ and 4 mJ. For a duty ratio of 30% in a DC drive application, determine the frequency at which the IGBTs A and B will give the same total power loss.

1.18 800-V IGBT, BJT, and MOSFET devices, of equal area, have on-state voltage drops of 1.5 V, 8 V, and 40 V, respectively, at a current of 25 A. The turn-off times of both the IGBT and the BJT are 2 μsec, while that of the MOSFET is 0.2 μsec. Find the power dissipation of these devices for a 60% duty cycle, if each device is carrying a current of 25 A. The operating frequency is 20 kHz.

REFERENCES

1. M. S. Adler, K. W. Owyang, B. J. Baliga, and R. A. Kokosa, The Evolution of Power Device Technology, *IEEE Trans. Electron Devices*, Vol. ED-31, No. 11, November 1984, pp. 1570–1591.
2. V. Rumennik, Power Devices are in the Chips, *IEEE Spectrum*, Vol. 22, No.7, July 1985, pp. 42–48.
3. B. J. Baliga, Evolution of MOS-Bipolar Power Semiconductor Technology, *Proc. IEEE*, Vol. 76, 1988, p. 409.
4. T. Mizogushi, T. Shirasawa, M. Mori, and Y. Sugawara, 600 V, 25 A Dielectrically-Isolated Power IC with Vertical IGBT, in *Proceeding of the 1991 IEEE International Symposium on Power Semiconductor Devices and ICs (ISPSD91)*, IEEE, New York, 1991, pp. 40–44.
5. B. K. Bose, Recent Advances in Power Electronics, *IEEE Trans. Power Electronics*, Vol. 7, No. 1, January 1992, pp. 2–16.

6. B. K. Bose (Ed.), *Modern Power Electronics-Evolution, Technology and Applications*, IEEE Press, New York, 1992.
7. S. M. Sze (Ed.), *Modern Semiconductor Device Physics*, John Wiley & Sons, New York, 1998.
8. R. Singh, S.-H. Ryu, and J. W. Palmour, High Temperature High Current, P-Channel UMOS 4H-SiC IGBT, in *57th Annual Device Research Conference Digest*, 1999, pp. 46–7.
9. L. Cao, B. Li and J. H. Zhao, Characterization of 4H-SiC Gate Turn-Off Thyristor, *Solid State Electronics*, Vol. 44, No 2, February 2000, pp. 347–352.
10. T. P. Chow, V. Khemka, J. Fedison, N. Ramungul, K. Matocha, Y. Tang and R. J. Gutmann, SiC and GaN Bipolar Power Devices, *Solid State Electronics*, Vol. 44, No. 4, February 2000, pp. 277–301.
11. P. K. Steimer, H. E Grüning, J. Werninger, E. Carroll, S. Klaka, and S. Lindner, IGCT-A New Emerging Technology for High Power, Low Cost Inverters, in *Conference Record of the IEEE Industry Applications Society (IEEE-IAS) Annual Meeting*, IEEE, New York, 1997, pp. 1592–1599.
12. S. Bernet, Recent Developments of High Power Converters for Industry and Traction Applications, *IEEE Trans. Power Electronics*, Vol. 15, No. 6, November 2000, pp. 1102–1117.
13. B. R. Pelly, Power MOSFETs: A Status Review, *Int. Power Electron. Conf. (IPEC) Rec.*, Vol. 1, March 1983, pp. 19–32.
14. (a) Coe, US Patent 4,754,310 (1988). (b) Chen, US Patent 5,216,275 (1993). (c) Tihanyi, US Patent 5,438,215 (1995).
15. L. Lorenz, Coolmos Technology-Outstanding Prospects Towards Idealized Power Semiconductor Switch, *European Power Electronics and Drives (EPE) Journal*, Vol. 10, No. 1, April 2000, pp. 11–16.
16. B. J. Baliga, Enhancement and Depletion Mode Vertical-channel MOS Gated Thyristors, *Electronics Lett.*, IEE, London, Vol. 15, No. 20, 27 September 1979, pp. 645–647.
17. J. D. Plummer and B. W. Scharf, Insulated Gate Planar Thyristors: I—Structure and Basic Operation, *IEEE Trans. Electron Devices*, Vol. ED-27, No. 2, February 1980, pp. 380–387.
18. L. Leipold, W. Baumgartner, W. Ledenhauf and J. P. Stengl, A FET-Controlled Thyristor in SIPMOS Technology, *International Electron Devices Meeting, IEDM Technical Digest*, IEEE, New York, 1980, p. 79.
19. J. Tihanyi, Functional Integration of Power MOS and Bipolar Devices, *International Electron Devices Meeting, IEDM Technical Digest*, IEEE, New York, 1980, p. 75.
20. H. W. Becke and C. F. Wheatley, Jr., *Power MOSFET with an Anode Region*, US Patent 4,364,073 (1982).
21. B. J. Baliga, M. S. Adler, P. V. Gray, R. P. Love and N. Zommer, The Insulated Gate Rectifier: A New Power Switching Device, *International Electron Devices Meeting, IEDM Technical Digest*, Abstract 10.6, 1982, pp. 264–267.
22. B. J. Baliga, M. Chang, P. Shafer and M. W. Smith, The Insulated Gate Transistor (IGT): A New Power Switching Device, *Conference Record of the IEEE Industry Applications Society (IEEE-IAS) Annual Meeting*, IEEE, New York, 1983, pp. 794-803.
23. J. P. Russell, A. M. Goodman, L. A. Goodman and J. M. Neilson, The COMFET: A New High Conductance MOS-gated Device, *IEEE Electron Device Lett.*, Vol. EDL-4, No. 3, March 1983, pp. 63–65.
24. M. F. Chang, G. C. Pifer, B. J. Baliga, M. S. Adler and P. V. Gray, 25 Amp, 500 Volt Insulated Gate Transistors, *International Electron Devices Meeting, IEDM Technical Digest*, IEEE, New York, Abstract 4.4, 1983, pp. 83–85.

25. A. M. Goodman, J. P. Russell, L. A. Goodman, C. J. Nuese and J. M. Neilson, Improved COMFETs with Fast Switching Speed and High-Current Capability, *International Electron Devices Meeting, IEDM Technical Digest*, IEEE, New York, Abstract 4.3, 1983, pp. 79–82.
26. B. J. Baliga, M. S. Adler, R. P. Love, P. V. Gray and N. D. Zommer, The Insulated Gate Transistor: A New Three Terminal MOS-Controlled Bipolar Power Device, *IEEE Trans. Electron Devices*, Vol. ED-31, No. 6, June 1984, pp. 821–828.
27. H. Yilmaz, W. R. Van Dell, K. Owyang, and M. F. Chang, Insulated Gate Transistor Modelling and Optimization, in *International Electron Devices Meeting, IEDM Technical Digest*, IEEE, New York, 1984, pp. 274–277.
28. A. Nakagawa, Y. Yamaguchi, K. Watanabe, H. Ohashi, and M. Kurata, Experimental and Numerical Study of Non-latchup Bipolar Mode MOSFET Characteristics, *International Electron Devices Meeting, IEDM Technical Digest*, IEEE, New York, Abstract 6.3, 1985, pp. 150–153.
29. M. Kitagawa, I. Omura, S. Hasegawa, T. Inoue, and A. Nakagawa, A 4500 V Injection Enhanced Insulated Gate Bipolar Transistor (IEGT) in a Mode Similar to a Thyristor, *International Electron Devices Meeting, IEDM Technical Digest*, IEEE, New York, 1993, pp. 679–682.
30. R. Zehringer, A. Stuck, and T. Lang, Material Requirements for High Voltage, High Power IGBT Devices, *Solid State Electronics*, Vol. 42, No. 12, December 1998, pp. 2139–2151.
31. D. S. Kuo, and C. Hu, An Analytical Model for the Power Bipolar-MOS Transistor, *Solid State Electronics*, Vol. 29, 1986. pp. 1229–1237.
32. J. G. Fossum, R. J. McDonald, and M. A. Shibib, Network Representations of LIGBT Structures for CAD of Power Integrated Circuits, *IEEE Trans. Electron Devices*, Vol. 35, No. 4, April, 1988, pp. 507–515.
33. A. R. Hefner and D. M. Diebolt, An Experimentally Verified IGBT Model Implemented in Saber Circuit Simulator, in *Record of the Annual IEEE Power Electronics Specialists Conference (PESC91), Proceedings PESC Record*, IEEE, New York, 1991, pp. 10–19.
34. R. Kraus and K. Hoffmann, Analytical Model of IGBTs with Low Emitter Efficiency, in *Proceedings, International Symposium on Power Semiconductor Devices and ICs (ISPSD93)*, IEEE, New York, 1993, pp. 30–34.
35. Z. Shen and R. P. Row, Modeling and Characterization of the Insulated Gate Bipolar Transistor (IGBT) for SPICE Simulation, in *Proceedings BN: IEEE International Symposium on Power Semiconductor Devices and ICs (ISPSD93)*, IEEE, New York, 1993, pp. 165–170.
36. S. M. Clemente and D. A. Dapkus, IGBT Models Account for Switching and Conduction Losses, *Power Conversion Intell. Motion*, Vol. 19, 1993, pp.51–54.
37. Y. Y. Tzou and I. J. Hsu, Practical SPICE Macro Model for the IGBT, *IECON (Ind. Electron. Conf.) Proc.*, Vol 2, The Annual International Conference of the IEEE Industrial Electronics Society, IEEE, New York, 1993, pp. 762–766.
38. F. F. Protiwa, O. Apeldoorn and N. Groos, New IGBT Model for PSPICE, *IEEE Conf. Publ.*, Vol. 2, 1993, pp. 226–231.
39. B. Fatemizadeh and D. Silber, A Versatile Electrical Model for IGBT Including Thermal Effects, *Proc. PESC Rec.—IEE Ann. Power Electron. Conf.*, 1993, pp. 85–92.
40. O. Kvein, T. M. Undeland, and T. Rogne, Models for Simulation of Diode (and IGBT) Switchings Which Include the Effect of the Depletion Layer, in *Conference Record of the IEEE Industry Applications Society Annual Meeting (IEEE-IAS)*, Vol 2, IEEE, New York, 1993, pp. 1190–1195.
41. J. M. Li, D. Lafore, J. Arnould, and B. Reymond, Analysis of Switching Behavior of the Power Insulated Gate Bipolar Transistor by Soft Modeling, in *Proc. 5th European*

Conference on Power Electronics and Applications, Brighton, U.K., EPE93, IEE, London, 1993, pp. 220–225.

42. J. B. Kuo and C. S. Chiang, Turn-on Transient Analysis of a Power IGBT with an Inductive Load in Series with a Resistive Load, *Solid State Electron,* Vol. 37, No. 9, September 1994, pp. 1673–1676.
43. V. A. Kuzmin, S. N. Yurkov, and L. I. Pomortseva, Analysis and Simulation of Insulated Gate Bipolar Transistor with Buffer N-Layer, *5th International Conference on Power Electronics and Variable Speed Drives,* 26–28 October, 1994, (PEVD-94), Conference Publication No. 399, IEE, London, pp. 24–28.
44. K. Besbes, Modeling an Insulated Gate Bipolar Transistor Using Bond Graph Techniques, *Int. J. Numer. Modeling: Electron. Networks, Devices Fields,* Vol. 8, 1995, pp. 51–60.
45. P. Spanik, B. Dobrucky, and R. Gubric, Dynamic Modeling of IGBT with Reverse Diode Modeling, *Meas. Contr.: General Phys., Electron., Electr. Eng.,* Vol. 59, 1995, pp. 23–32.
46. F. Mihalic, K. Jezernik, D. Krischan, and M. Rentmeister, IGBT SPICE Model, *IEEE Trans. Ind. Electron.,* Vol. 42, No. 1, February 1995, pp. 98–105.
47. C. Alonso and T. A. Meynard, Simulation of Short-Circuit Phenomena in IGBT, *Proc. IEE Colloq. (Dig.),* 1994, pp. 10/1–5.
48. A. R. Hefner, Modeling Buffer Layer IGBTs for Circuit Simulation, *IEEE Trans. Power Electronics,* Vol. 10, 1995, pp. 111–123.
49. C. Wong, EMTP Modeling of IGBT Dynamic Performance for Power Dissipation Estimation, in *Conference Record of the IEEE Industry Applications Society (IEEE-IAS) Annual Meeting,* Vol. 3, 1995, IEEE, New York, pp. 2656–2662.
50. F. Udrea and G. A. J. Amaratunga, A Unified Model for Carrier Dynamics in Trench Insulated Gate Bipolar Transistor (TIGBT), in *Proceedings IEEE International Symposium on Power Semiconductor Devices & ICs (ISPSD),* IEEE, New York, 1995, pp. 190–200.
51. F. Udrea and G. A. J. Amaratunga, Steady-State Analytical Model for the Trench Insulated Gate Bipolar Transistor, in *Proc. Int. Semiconductor Conf., CAS,* 1995, pp. 49–52.
52. F. Udrea and G. A. J. Amaratunga, An On-State Analytical Model for the Trench Insulated Gate Bipolar Transistors (TIGBT), *Solid State Electronics,* Vol. 4, No. 8, August, 1997, pp. 1111–1118.
53. A. G. M. Strollo, New IGBT Circuit Model for SPICE Simulation, in *Record of the Annual IEEE Power Electronics Specialists Conference (PESC97),* Vol. 1, IEEE, New York, 1997, pp. 133–138.
54. T. Yamazaki,Y. Seki,Y. Hoshi, and N. Kumagai, The IGBT with Monolithic Overvoltage Protection Circuit, in *Proc. IEEE International Symposium on Power Semiconductor Devices and ICs (ISPSD93),* IEEE, New York, 1993, pp. 41–45.
55. Y. Seki, Y. Harada, N. Iwamuro, and M. Kumagai, A New IGBT with a Monolithic Overcurrent Protection Circuit, in *Proceedings 6th International Symposium on Power Semiconductor Devices and ICs (ISPSD94)* IEEE, New York, 1994, pp. 31–35.
56. A. Pshaenich, The MOS-SCR, A New Thyristor Technology, *Motorola Engineering Bulletin,* ED-103, 1982.
57. V. A. K. Temple, *MOS Controlled Thyristors (MCTs), International Electron Devices Meeting, IEDM Technical Digest,* IEEE, New York, Abstract 10.7, 1984, pp. 282–285.
58. M. Stoisiek and H. Strack, MOS-GTO: A Turn-Off Thyristor with MOS-Controlled Emitter Shorts, in *International Electron Devices Meeting, IEDM Technical Digest,* IEEE, New York, 1985, pp. 158–161.
59. V. A. K. Temple, MOS-Controlled Thyristors—A New Class of Power Devices, *IEEE Trans. Electron Devices,* Vol. ED-33, No. 10, October 1986, pp. 1609–1618.

60. V. A. K. Temple, S. Arthur and D. L. Watsons, MCT (MOS Controlled Thyristor) Reliability Investigation, 1988, *International Electron Devices Meeting, IEDM Technical Digest*, IEEE, New York, pp. 618–621.
61. S. K. Sul, F. Profumo, G. H. Cho, and T. A. Lipo, MCTs and IGBTs: A Comparison of Performance in Power Electronic Circuits, *Proceedings Record of the Annual IEEE Power Electronics Specialists Conference (PESC89)*, IEEE, New York, 1989, pp. 163–169.
62. F. Bauer, T. Roggwiler, A. Aemmer, W. Fichtner, R. Vuilleumier, and J. M. Moret, Design Aspects of MOS Controlled Thyristor Elements, *International Electron Devices Meeting, IEDM Technical Digest*, IEEE, New York, 1989, p. 297.
63. T. M. Jahns, R. W. A. A. DeDoncker, J. W. A. Wilson, V. A. K. Temple, and D. L. Watrous, Circuit Utilization Characteristics of MOS-Controlled Thyristor, *EEE Trans. Ind. Appl.*, Vol. 27, No. 3, May/June 1991, pp. 589–597.
64. Q. Huang, G. A. J. Amaratunga, E. M. S. Narayanan, and W. I. Milne, Analysis of N-Channel MOS-Controlled Thyristors, *IEEE Trans. Electron Devices*, Vol. ED-38, No. 7, July, 1991, pp. 1612–1618.
65. F. Bauer, T. Stockmeier, H. Lendenmann, H. Dettmer and W. Fichtner, Static and Dynamic Characteristics of High Voltage (3 kV) IGBT and MCT, in *IEEE International Symposium on Power Semiconductor Devices and ICs (ISPSD92)*, IEEE, New York, 1992, p. 22.
66. R. W. A. A. De Doncker, T. M. Jahns, A. V Radun, D. L. Watrous, and V. A. K. Temple, Characteristics of MOS-Controlled Thyristors Under Zero Voltage Soft Switching Conditions, *IEEE Trans. Ind. Appl.*, Vol. 28, No. 2, March/April 1992, pp. 387–394.
67. H. Dettmer, W. Fichtner and F. Bauer, New (Super) Self-Aligned MOS-Controlled Thyristors, (S) SAMCTs, in *International Electron Devices Meeting, IEDM Technical Digest*, IEEE, New York, 1994, pp. 395–398.
68. E. R. Motto and M. Yamamoto, New High Power Semiconductors: High Voltage IGBTs and GCTs, in *Proc. PCIM98 Power Electron. Conf.*, 1998, pp. 296–302.
69. S. B. Bayne, W. M. Portnoy, and A. R. Hefner, Jr., MOS-Gated Thyristors (MCTs) for Repetitive High Power Switching, *IEEE Trans. Power Electronics*, Vol. 16, No. 1, January 2001, pp. 125–131.

2

IGBT FUNDAMENTALS AND STATUS REVIEW

In this chapter the basics of IGBT operation will be reviewed. The chapter will describe the various categories of IGBTs, their operating modes, electrical and thermal characteristics, radiation effects, and so on.

2.1 DEVICE STRUCTURES

2.1.1 Lateral IGBT (LIGBT) and Vertical IGBT (VIGBT)

Lateral IGBT or LIGBT [1–9] lends itself more readily to integration on the same chip thus helping in the realization of power integrated circuits (PICs) and smart power ICs. It is mainly used for integration of power device with control circuitry. The lateral approach consumes a great deal of Si surface area for supporting the voltage and therefore furnishes a low silicon utilization factor. Moreover, latchup problems (due to the inability to use lifetime killing techniques in an IC environment) and substrate current are causes for concern. For these reasons, it is commonly avoided in discrete devices where the vertical structure described in previous chapter is commonly used.

As shown in the cross-sectional view of the lateral IGBT (Fig. 2.1a), the collector contact is taken from the top surface of the chip instead of the bottom surface in the vertical structure. In the LIGBT, a P^+ collector replaces the N^+ drain of the LDMOS. Unlike the vertical structure, there is no JFET region in the lateral IGBT. Furthermore, during ON-state operation, the accumulation region buildup under the gate exerts a strong influ-

---

The Insulated Gate Bipolar Transistor (IGBT) Theory and Design, By Vinod Kumar Khanna.
ISBN 0-471-23845-7 © 2003 Institute of Electrical and Electronics Engineers

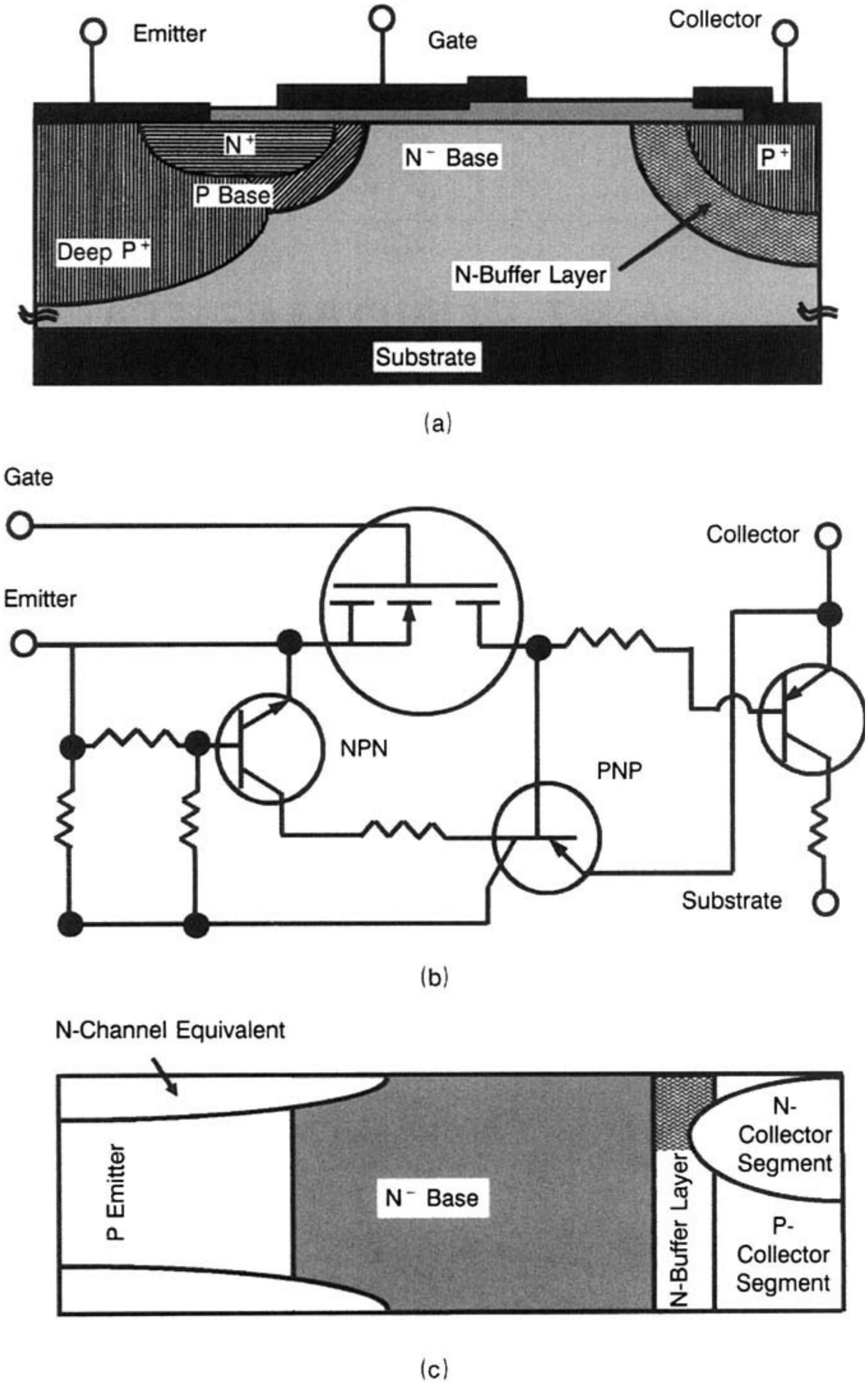

Emitter
Gate
Collector
N+
N- Base
P+
P Base
Deep P+
N-Buffer Layer
Substrate
(a)
Gate
Collector
Emitter
NPN
PNP
Substrate
(b)
N-Channel Equivalent
P Emitter
N- Base
N-Buffer Layer
N- Collector Segment
P- Collector Segment
(c)

Figure 2.1 Lateral IGBTs. (a) Cross-sectional view of lateral IGBT. (b) Equivalent circuit of lateral IGBT. (c) In-plane diagram of shorted collector IGBT half-cell.

ence on the carrier distribution in the drift region. Although an LIGBT conducts current laterally like LDMOS, the presence of back-to-back PN junctions enables it to block voltage in both directions. A modified latchup-resistant LIGBT structure [3], containing an N^+ sinker for diverting hole flow away from the P-base toward an auxiliary emitter and also containing a P^+ buried layer, is more suitable at high operating temperatures.

Figure 2.1b presents the electrical equivalent circuit of an LIGT. It is the combination of a lateral DMOSFET, a vertical BJT, a lateral BJT, and a lateral parasitic thyristor. Resistances of the semiconductor paths interconnecting these components are also included. On biasing the DMOSFET gate above the threshold voltage, its source serves as the base electrode of both the lateral and vertical BJTs. Hence this base current increases with gate-source voltage, accompanied by rise of emitter current by transistor action.

Incorporation of N^+ segments into the collector, penetrating up to the N-buffer layer, improves the turn-off behavior of LIGBT [4–8]. This IGBT is called *segmented collector IGBT* (SC-IGBT) or *shorted collector IGBT*. In-plane view of the half-cell of this IGBT cell containing one P-collector segment is shown in Fig. 2.1c. Technologically, with regard to the planar IC process, this method of controlling the injection efficiency of the P^+ collector

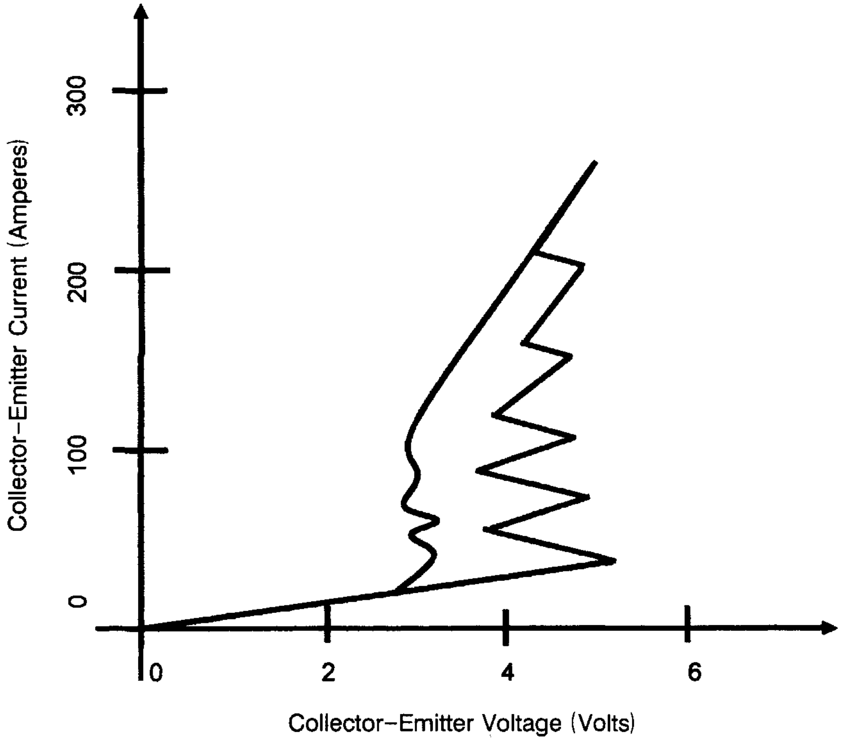


Figure 2.2 Voltage-controlled multistability and hysteresis in IGBT current–voltage characteristics.

is superior to carrier lifetime monitoring. The current–voltage characteristics of these IGBTs exhibit S-type negative differential conductivity (SNDC). Certain collector designs show voltage-controlled multistability and hysteresis in stationary characteristics (Fig. 2.2). To understand the meaning of *multistability* in characteristics, it may be noted that as the collector–emitter voltage increases, the collector–emitter current rises until at a certain voltage the voltage falls back to a smaller value at slightly higher current. Thereafter, further increase of collector–emitter voltage is accompanied by increase of current. This is again followed by the decline of voltage at somewhat higher current. The above sequence of events may be repeated several times. Instabilities arise from the switching of the single P-emitter segments between the MOS mode and the conductivity-modulated mode and vice versa. *Hysteresis* indicates that the sweep-down branch of the characteristic does not retrace the sweep-up branch. Physically, hysteresis is ascribed to the presence of two or more stable stationary states. For small disturbances the system prefers to retain its original state, but perturbations such as changes in external voltage or current cause the state to lose its stability parameter value, leading to hysteresis.

2.1.2 Non-punchthrough IGBT (NPT-IGBT) and Punchthrough IGBT (PT-IGBT)

After discussing the lateral IGBT structure, we are now transitioning to the vertical structures. Commercially available IGBTs are broadly classified un-

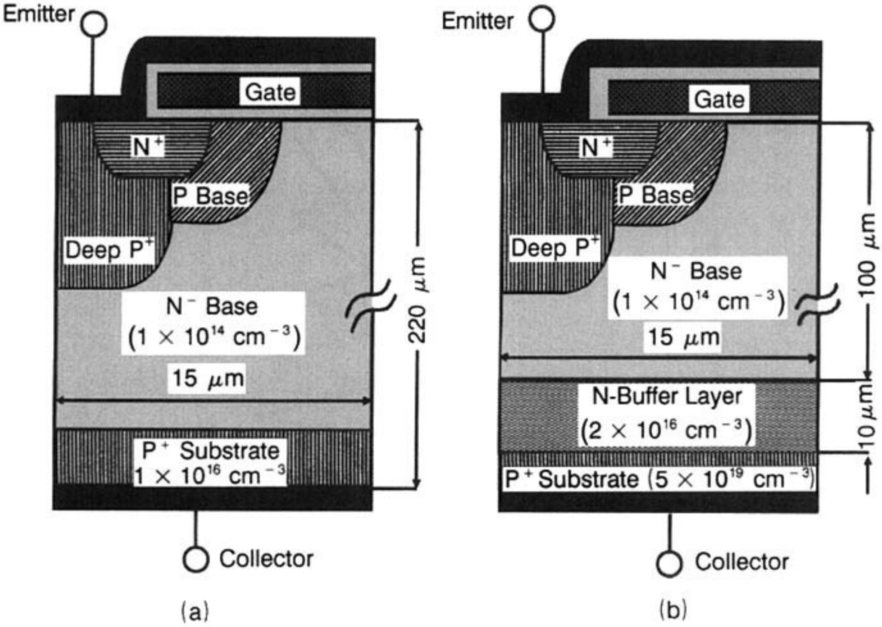

Emitter
Gate
N+
P Base
Deep P+
N- Base
(1 × 10^14 cm^-3)
15 μm
220 μm
P+ Substrate
1 × 10^16 cm^-3
Collector
(a)
Emitter
Gate
N+
P Base
Deep P+
N- Base
(1 × 10^14 cm^-3)
15 μm
100 μm
N-Buffer Layer
(2 × 10^16 cm^-3)
10 μm
P+ Substrate (5 × 10^19 cm^-3)
Collector
(b)

Figure 2.3 Two IGBT structures: (a) NPT-IGBT and (b) PT-IGBT.

Table 2.1 NPT-IGBT Versus PT-IGBT

Serial No.	Feature	NPT-IGBT	PT-IGBT
1	Process technology and cost-effectiveness	Manufactured using diffusion steps. Less expensive process.	Fabricated in an N^- epitaxial wafer. More expensive.
2	N-buffer layer and N^--base thickness	Thick N base. Does not contain any N-buffer layer. Space charge spreads across the wide N^- base to withstand the voltage. NPT structure provides bidirectional blocking capability.	Thin N base. Contains an N-buffer layer. Penetration of depletion region into this layer avoids the use of a broad N^- base. This IGBT has lower reverse blocking capability.
3	Carrier lifetime in N^- base and conductivity modulation	High carrier lifetime yields a low forward drop.	Lower lifetime able to provide adequate conductivity modulation as the N base is thin. Forward drop is determined by the carrier lifetime in N^- base and injection efficiency of P^+ substrate.
4	Temperature coefficient of forward voltage drop; paralleling	The temperature coefficient of ON-state voltage is strongly positive. Therefore, simple paralleling.	Small positive temperature coefficient of ON-state voltage. Paralleling calls for greater care and attention.
5	Collector doping and turn-off time	Collector is lightly doped (P only). Electron back injection from N^- base into P collector gives satisfactory turn-off time.	Heavy doped collector (P^+). Lifetime killing in N base is necessary to achieve the required turn-off time. Injection efficiency reduction of the P^+ substrate by the buffer layer makes its fall time and the current tail shorter.
6	Turn-off loss	Turn-off loss is less temperature-sensitive, and it remains practically unchanged with temperature.	Turn-off loss is more temperature-sensitive, and it increases significantly at higher temperature.
7	Thermal stability	More thermally stable.	Less thermally stable. Thermal run-away occurs at a lower junction temperature.
8	Short-circuit failure	More rugged in short-circuit failure mode.	Less rugged.

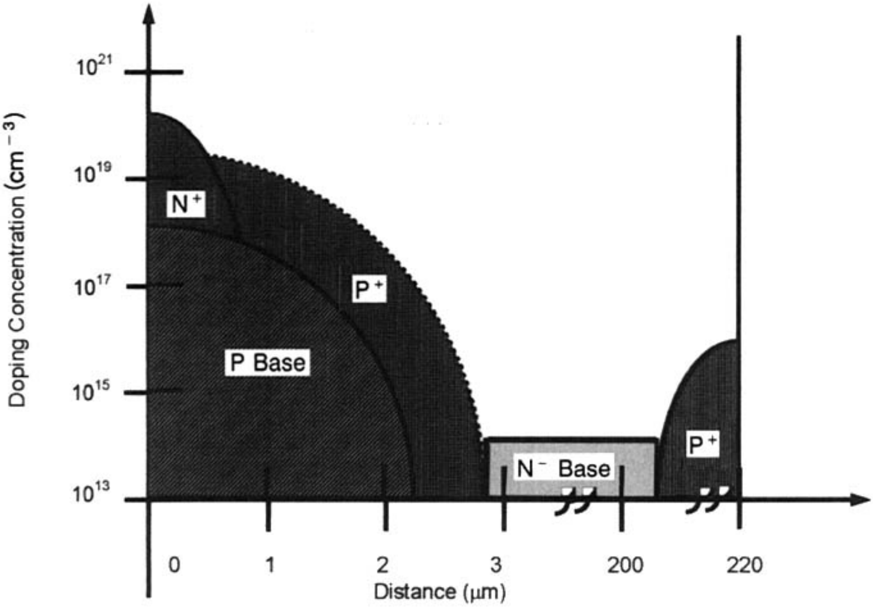
der two headings: NPT-IGBT and PT-IGBT [10–18]. These IGBTs are also referred to as *symmetrical* and *asymmetrical* IGBTs. These varieties of IGBT, shown in Fig. 2.3, differ widely with regard to their fabrication technology, structural details, carrier profiles, lifetimes, and transport mechanisms. Table 2.1 looks at the salient features of the two types of IGBTs. The constructional and operational demarcation translates into the different output characteristics, behavioral pattern, and degree of thermal ruggedness, observed for these devices.

NPT-IGBT fabrication is based on thin wafer technology. The starting Si wafer is an N-type ⟨100⟩ orientation wafer of thickness 220 μm and concentration 1×10^{14} cm^{-3} for the 1200-V device (see Fig. 2.3a). On the backside of this wafer, boron implantation is performed followed by a long, high-temperature, drive-in cycle. This produces a shallow doped P-type collector of concentration $\sim 10^{16}$ cm^{-3}, instead of P$^+$ collector (Fig. 2.4a). Obviously, the concentration gradient across the P-type collector/N$^-$ base is low. As a consequence, the carrier lifetime distribution in the N$^-$ base is uniform so that the carrier flow is not much aided by diffusion but drift-controlled. Furthermore, the electric field distribution across the N$^-$-base thickness makes the device sturdier.

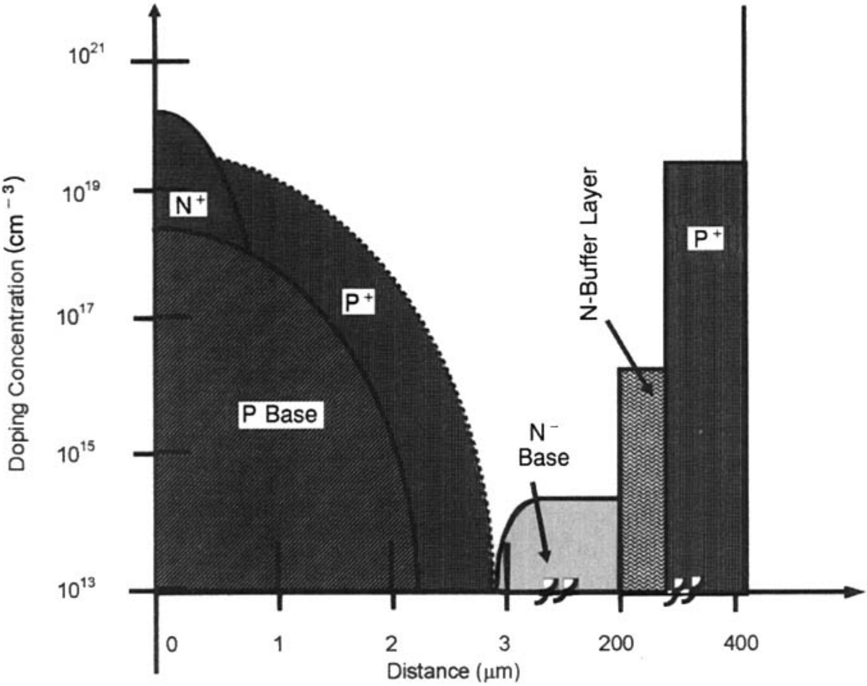
PT-IGBT fabrication (Fig. 2.3b) utilizes the technology of growing a thick epitaxial layer of uniform resistivity (concentration $= 1 \times 10^{14}$ cm^{-3}) and thickness (100 μm) on P$^+$ substrate (5×10^{19} cm^{-3}). The thick epitaxial layer forms the N$^-$ base of the IGBT. Sandwiched between the N$^-$ base and P$^+$ collector is an N-buffer layer (2×10^{16} cm^{-3}) of thickness 10 μm (Fig. 2.4b). This buffer layer has twofold functions. First, by using the buffer layer, the thickness of the N$^-$ base required to support a given breakdown voltage is decreased, since the depletion layer expansion at high applied voltage is accommodated in a small thickness of this layer, avoiding failure by punch-through. Second, the holes injected by the P$^+$ collector partially recombine in the buffer layer. This slightly offsets the injection efficiency of the collector, preventing it from becoming too high to degrade the turn-off behavior of IGBT. Thus the buffer layer reduces the tail current during turn-off and shortens the fall time of IGBT. By this mechanism, the trade-off between conduction and switching losses is improved.

Due to the existence of a high concentration gradient across the P$^+$ collector/N-buffer layer/N$^-$ base, the current flow takes place primarily by diffusion in the PT-IGBT as compared to the drift transport in an NPT-IGBT. Furthermore, because the electric field has a smaller N$^-$-base thickness to spread over, the PT device becomes more fragile than the NPT case.

Now let us inquire as to what makes the NPT-IGBT and PT-IGBT "symmetrical" and "asymmetrical" in behavior. A *symmetrical* IGBT (the NPT-IGBT) is one having equal forward and reverse breakdown voltages. Such a device is used in AC applications. In the *asymmetrical* or PT-IGBT structure, the reverse breakdown voltage is less than the forward breakdown voltage. This IGBT is useful for DC circuits where the device is not required to support voltage in the reverse direction. The notable advantage of the



(a)



(b)

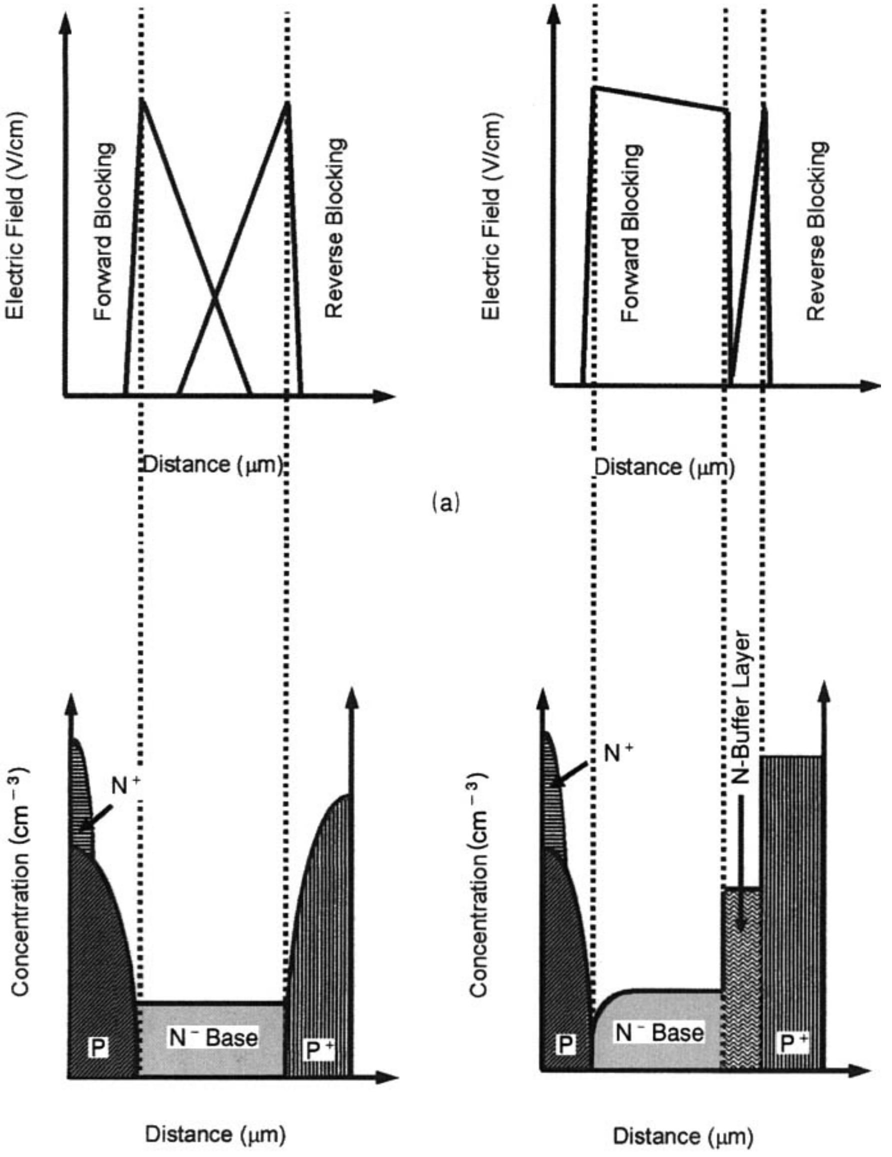
Figure 2.4 Representative impurity diffusion profiles of (a) NPT-IGBT and (b) PT-IGBT.

asymmetrical structure is the provision of opportunity for improvement of forward conduction characteristics by sacrificing the reverse blocking capability. As explained above, the common method of accomplishing this is by employing a double-layer base structure consisting of an N^- layer and an N^+ buffer layer. It has also been mentioned that the buffer layer allows us to use a thinner N^- base producing a comparatively smaller forward drop.

The doping profiles of the symmetrical and asymmetrical IGBTs along with the resulting electric field distributions are illustrated in Fig. 2.5. If the critical breakdown field is assumed to be independent of the doping level, the shape of the electric field distribution changes from *triangular* in the symmetrical IGBT to *trapezoidal* for the asymmetrical IGBT. When the buffer layer thickness equals one diffusion length, the forward breakdown voltage of asymmetrical IGBT will be twice that of the symmetrical IGBT. Accounting for the decrease in maximum electric field with base width reduction due to field redistribution over a larger space together with the finite N^--base doping required for design optimization, the increase in breakdown voltage is achieved by a factor between 1.5 and 2.0.

During design of the impurity diffusion profile of asymmetrical IGBT, it is desirable that the buffer layer thickness be kept minimum, consistent with the required blocking voltage for making a fast switching device. This is only possible by raising the doping concentration of this layer to a level at which punchthrough to the N-buffer layer/P^+ junction is avoided. But when the doping concentration of the buffer layer becomes too high, the injection efficiency of P^+ layer is crippled, lowering the current gain of the PNP transistor. This has an adverse influence on the forward drop of IGBT but has the favorable impact of increasing the collector output resistance.

Let us understand the effect of buffer layer on the collector output resistance. Figure 2.6 illustrates pictorially the variation of depletion layer width of symmetric and asymmetric IGBT structures with increasing collector–emitter voltage (V_{CE}). It is evident that in the symmetrical IGBT, the depletion layer width continuously increases with collector–emitter voltage. Consequently, the undepleted N^--base width (W) narrows down, producing an increase in the current gain of the PNP transistor $\alpha_{PNP} = 1/\{\cosh(W/L_p)\}$ [where L_p is the minority-carrier (hole) diffusion length in N^- base] and thereby in the output collector–emitter current (I_{CE}). The overall impact is that the I_{CE}–V_{CE} characteristics of an asymmetric IGBT show a steady increase in I_{CE} with V_{CE}, instead of saturating like a power MOSFET. In a power MOSFET, this upward trend is only caused by the decrease in channel length with V_{CE} , but in the symmetric IGBT the undepleted base width constriction is an additional feature. Therefore, the collector output resistance of a symmetric IGBT defined as the reciprocal slope of the I_{CE}–V_{CE} curve (i.e., $r_c = \partial V_{CE}/\partial I_{CE}$) is less than that for the equivalent MOSFET. The same is not true for the asymmetrical IGBT whose undepleted N^--base width remains constant at the buffer layer thickness for all the values of V_{CE}. So, the collector output resistance of the asymmetrical IGBT is comparable to that of the equivalent MOSFET and higher than that in the symmetrical


Electric Field (V/cm)
Forward Blocking
Reverse Blocking
Distance (μm)
Electric Field (V/cm)
Forward Blocking
Reverse Blocking
Distance (μm)
Concentration (cm⁻³)
N⁺
P
N⁻ Base
P⁺
Distance (μm)
Concentration (cm⁻³)
N⁺
N-Buffer Layer
P
N⁻ Base
P⁺
Distance (μm)

(a)

(b)

Figure 2.5 Doping profiles and electric field distribution in (a) NPT-IGBT (symmetrical) and (b) PT-IGBT (asymmetrical).

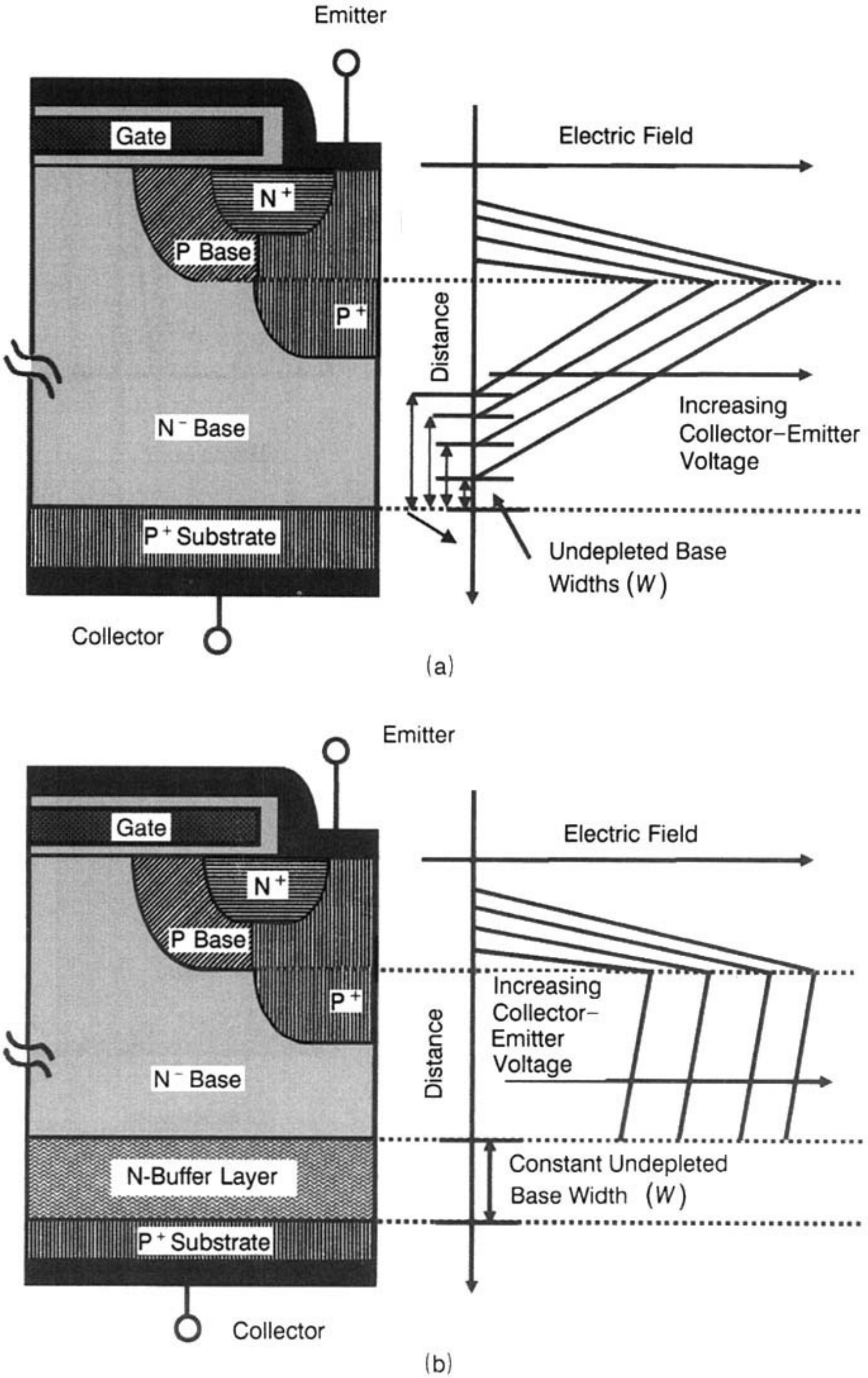

Emitter
Gate
N+
P Base
P+
N- Base
P+ Substrate
Collector
Electric Field
Distance
Increasing Collector–Emitter Voltage
Undepleted Base Widths (W)
(a)
Emitter
Gate
N+
P Base
P+
N- Base
N-Buffer Layer
P+ Substrate
Collector
Electric Field
Distance
Increasing Collector–Emitter Voltage
Constant Undepleted Base Width (W)
(b)

Figure 2.6 Depletion layer spreading in (a) NPT-IGBT (symmetrical) and (b) PT-IGBT (asymmetrical).

case. This will be elaborated later on. Based on the above discussions, a compromise solution is sought, striking a balance between forward drop, switching speed, and collector output resistance. ON-state voltage drop and turn-off time trade-off analysis shows that the asymmetric IGBT yields lower power losses than does the symmetric IGBT.

2.1.3 Complementary Devices

Complementary devices like N-channel and P-channel MOSFETs are required in power conditioning circuits such as variable-speed motor drives. Examination of the performance of N-channel and P-channel MOSFETs having the same chip area, epilayer resistivity, and thickness reveals that the ON resistance of the N-channel MOSFET is far less than that of the P-channel MOSFET. This is primarily due to the much larger mobility of electrons than holes, accounting for the higher resistance of the P-epitaxial layer as well as the higher channel resistance in the P-channel case. Specifically, because the electron mobility is three times the hole mobility, the P-channel MOSFET must be made three times larger in area to be able to handle the same power as an N-channel device.

The question may be posed, "Does the same hold true for N-channel and P-channel IGBTs?" Fortunately, the answer is "No," so that IGBT designers do not face such a problem. N- and P-channel IGBTs of a particular die size and breakdown voltage rating exhibit identical forward voltage drops so that their current-carrying capabilities are similar [19, 31, 33]. The reason is that the IGBT ON resistance is the sum total of contributions from channel resistance and drift region resistance. For low forward drop, slow IGBTs, carrier lifetime in the drift region is quite high. So, the governing PNP transistor in an N-channel IGBT and, likewise, the dominant NPN transistor in a P-channel IGBT have large current gains, producing adequate conductivity modulation of the drift region. Hence, the predominant component of ON resistance, arising from the epitaxial drift layers, is approximately the same for the two kinds of IGBTs, being essentially governed by conductivity modulation of this region. This modulation effect overshadows any difference of resistivity of epitaxial layers. However, for the faster, higher forward voltage IGBTs, the lifetime of carriers in the drift region is low. As a result, considering IGBTs of equal voltage ratings, the current gain of the PNP transistor in N-channel IGBT is small and so, also is the gain of NPN transistor in P-channel IGBT. Conductivity modulation is therefore, not as effective as in the slow devices. Now channel resistance cannot be ignored and their contributions become decisive. So, the difference in ON resistance of P-channel and N-channel IGBTs becomes more pronounced. Even in this situation, the superiority of IGBTs over MOSFETs is maintained, because the IGBTs provide lower forward drops under certain operating conditions. In the limiting case, when the lifetime has been killed to the extent that the channel resistance component begins to dominate in the ON resistance,

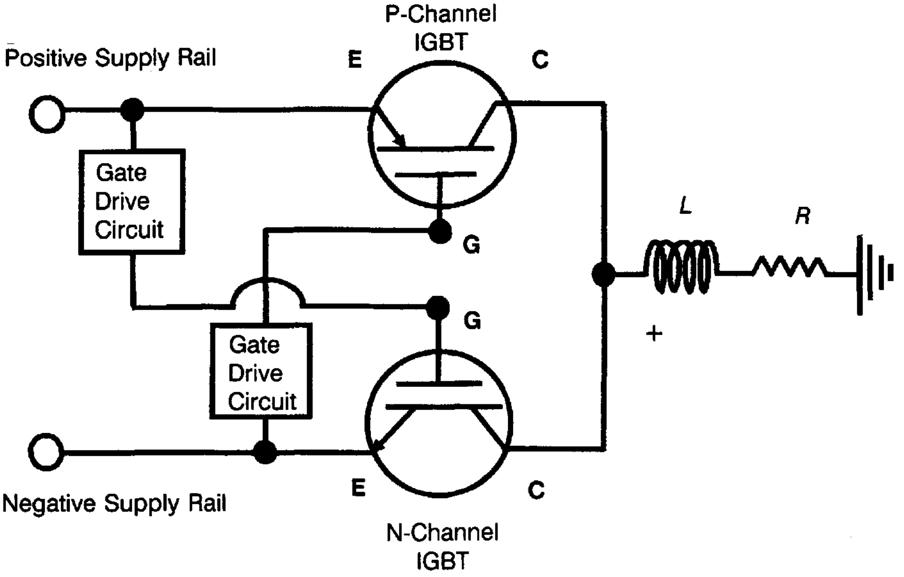


Figure 2.7 Complementary half-bridge inverter circuit.

the P-channel IGBT has a forward drop that is three fold larger than that of the N-channel device. This may occur, for example, for very fast devices.

The advantage of availability of complementary IGBT devices is utilized in the realization of high-voltage bridge circuits. Here, both the IGBTs function with the emitter connected to one supply rail and the gate voltage obtained from the other supply rail (Fig. 2.7). This feature offered by IGBTs makes the design of bridge circuits easier. Other applications involving N- and P-channel IGBTs include numerical and appliance controls where an N-channel IGBT is connected in parallel with a P-channel IGBT constituting a composite AC switch which allows the controlling of both the IGBTs using a common reference terminal.

Here, it must be pointed out that a comparison of switching times shows that P-channel IGBTs are 2–3 times slower than N-channel IGBTs of the same forward drop. Also, due to a smaller SOA, the maximum controllable current during high-voltage switching is lower for the P-channel variety.

2.2 DEVICE OPERATIONAL MODES

Device operation can be easily understood with reference to the different operating modes of the NPT-IGBT as follows:

2.2.1 Reverse-Blocking Mode

With the gate terminal shorted to the emitter terminal, a positive bias is applied to the N^+ emitter while a negative bias is applied to the P^+ collector

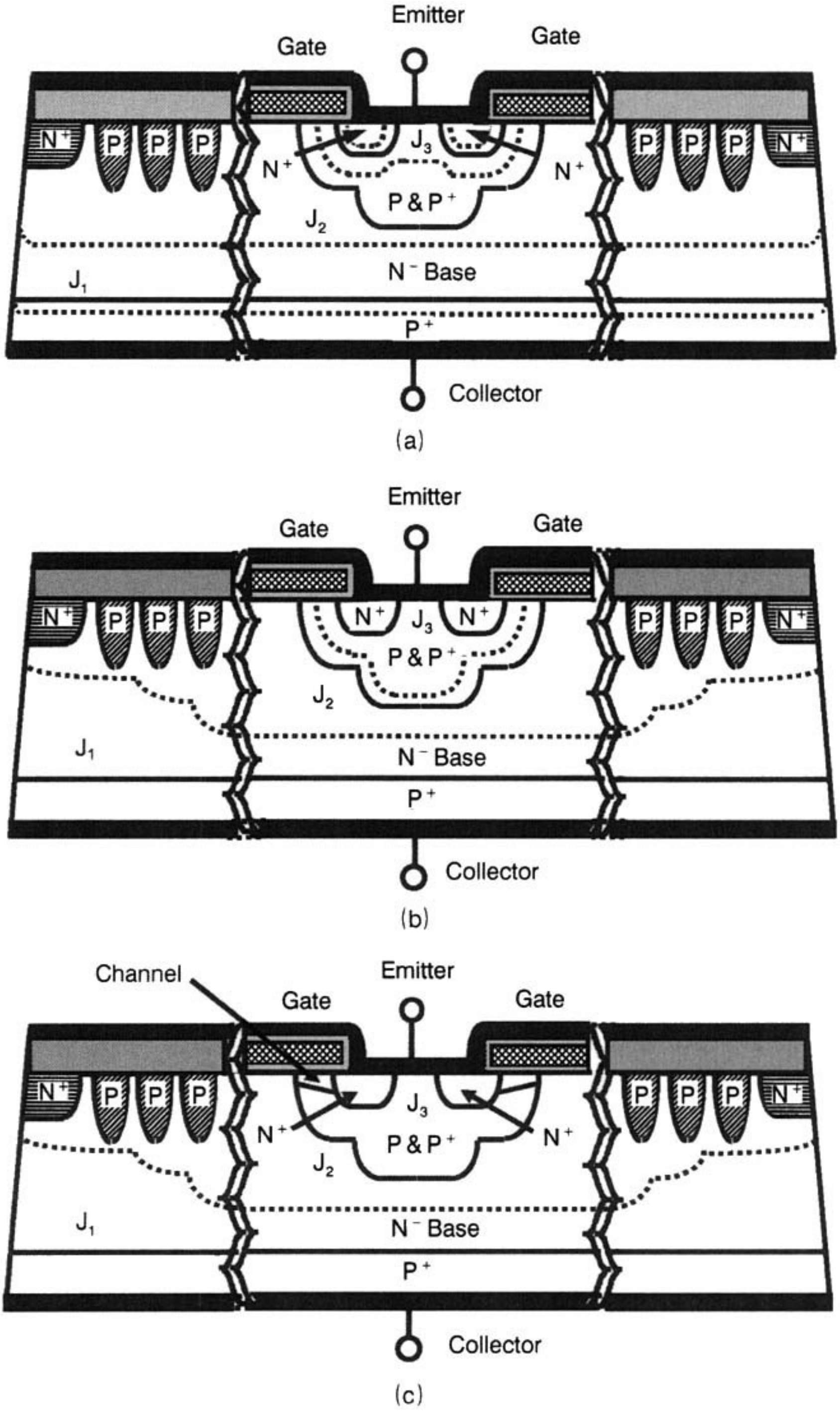

Emitter
Gate
Gate
N+
P
P
P
N+
J3
N+
N+
P & P+
J2
J1
N- Base
P+
Collector
(a)
Emitter
Gate
Gate
N+
P
P
P
N+
J3
P & P+
J2
J1
N- Base
P+
Collector
(b)
Channel
Emitter
Gate
Gate
N+
P
P
P
J3
N+
N+
P & P+
J2
J1
N- Base
P+
Collector
(c)

Figure 2.8 Operational modes of IGBT. (a) Reverse blocking mode. (b) Forward blocking mode. (c) Forward conducting state.

(Fig. 2.8a). In this condition, junctions J_1 and J_3 are reverse-biased while the junction J_2 is forward-biased. The reverse-biased junctions, J_1 and J_3, inhibit the current flow through the device, imparting to it the *reverse-blocking capability* [20, 21]. The voltage is supported mainly across the junction J_1. A large fraction of the depletion region extends across the low-doped N^- base and slightly across the P^+ collector due to its heavy doping.

The reverse blocking voltage is determined by an open-base transistor comprising P^+ collector, N^--base and P-base regions. This transistor may suffer breakdown by reach-through for the thin N^- base, or light doping of this region. The resistivity and thickness of the N^- base must therefore be optimized for the required breakdown voltage. A rule of thumb is to choose the N^--base thickness (d) equal to the depletion layer width at the maximum operating voltage ($V_{\max}$), plus one minority-carrier diffusion length (L_p), that is

$$d = \sqrt{\frac{2\varepsilon_0 \varepsilon_s V_{\max}}{qN_D}} + L_p \tag{2.1}$$

where ε_0 is permittivity of free space, ε_s is dielectric constant of Si, q is the electronic charge, and N_D is the doping concentration of N^- base. At high blocking voltages, the first term dominates and the N^--base thickness $d \propto \sqrt{V_{\max}}$.

Angle beveling, at the edges of the chip, prevents premature surface breakdown. This is accomplished during wafer dicing for chip separation. A tapered tool is used for dicing so that the chip is shaped in the form of a mesa or plateau, broader on the P^+ collector side and narrower at the opposite end. As the junction area increases, upon moving from the lower doped N^- base to the higher doped P^+ collector, a positive bevel angle is obtained. Obviously, the bevel angle control is not as stringent here as it is when the bevel angle is negative.

The IGBT structure is particularly suited to high-voltage applications. Performance of 300-, 600-, and 1200-V IGBTs has been experimentally compared [20]. These studies have shown that the forward current-carrying capability is, approximately, inversely proportional to $\sqrt{\text{breakdown voltage}}$. The slow fall of current-carrying capacity with increase in breakdown voltage has resulted in the realization of high-current and high-voltage IGBTs (e.g., 1000-A, 2500-V modules).

2.2.2 Forward-Blocking and Conduction Modes

Upon reversing the emitter and collector supply voltages with the gate shorted to emitter as before, junctions J_1 and J_3 are forward-biased but junction J_2 is reverse-biased, withstanding the voltage across itself (Fig.2.8b). *Forward-blocking capability* is thus obtained in the IGBT [20]. The depletion layer stretches partly into the P base. Most of it expands into the N^- base. At

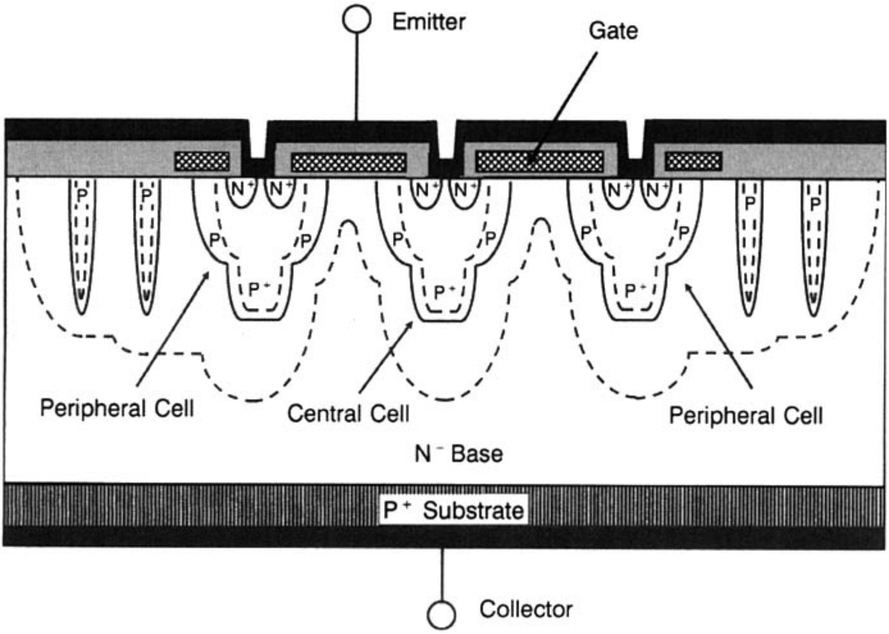


Figure 2.9 Spreading of the depletion layer in the central and peripheral regions of an IGBT device.

higher applied voltages, the depletion region of each cell merges with that of the neighboring cells on all the sides (Fig. 2.9). At the boundaries of the chip, optimally spaced field rings subdivide the voltage, shape, smooth the depletion region and safely terminate it at the surface.

An IGBT in forward-blocking state can be transferred to the forward conducting state by removing the gate–emitter shorting and applying a positive voltage of sufficient level to invert the Si below in the P base (Fig. 2.8c). It must be noted that under this condition the supply voltages on the three terminals are as follows: collector-positive, emitter-negative, and gate-positive.

The N-type conducting channel, formed in the P-base region, bridges across the N^+ emitter and N^- base. Through this channel electrons are transported from the N^+ emitter to the N^--base. This flow of electrons into the N^--base region lowers the potential of the N^- region whereby P^+ collector/N^--base diode becomes forward-biased. Under this forward bias, a high density of minority carrier holes is injected into the N^- base from the P^+ collector. At high values of collector voltage, the injected hole concentration may be about 10^2 to 10^3 times the background doping of N^- base region. When the injected carrier concentration is very much larger than the background concentration, a condition of *high-level injection* is said to prevail in the N^- region of the IGBT. A hole injected by the P^+ collector from a point directly below the emitter-base short traverses a rectilinear path through the N^- base/P base and reaches the N^--base/P-base space-charge region. Finally it arrives at the emitter contact after crossing the P^+ region. Another hole, which is ejected from a point directly below the gate region, travels vertically upwards and reaches under the gate. Here it is repelled by the positively charged *accumulation layer* below the gate and diverted toward the P-base/N^--base space-charge region. Moving underneath the N^+ emitter and through the P^+ region, it is captured by the emitter contact. Thus at high collector supply voltages, a *plasma of holes* builds up in the N^- base. This plasma of holes attracts electrons from the emitter contact to maintain local charge neutrality. In this manner, approximately equal excess concentrations of holes and electrons are gathered in the N^- base. These excess electron and hole concentrations drastically enhance the conductivity of N^- base. The mechanism of rise in conductivity is referred to as *conductivity modulation* of N^- base.

Now focusing our attention on the electronic current flow, the electrons travel from the N^+ emitter, through the N channel, the N^- base, and eventually end up at the P^+ base. As conductivity modulation of the N^- base has transformed it into a high conductance region, electron flow through this region does not encounter much opposition. The ON resistance and thereby the forward voltage drop is significantly reduced.

From the above discussions, an IGBT in its forward conduction state can be visualized as a P-I-N rectifier whose output current is constrained to flow through a MOSFET channel. This permits the control of P-I-N diode current by an applied voltage. On this basis, the IGBT can be viewed as a series connection of MOSFET and P-I-N rectifier. This provides an equivalent circuit model for the IGBT to be discussed in Chapter 5.

2.3 STATIC CHARACTERISTICS OF IGBT

2.3.1 Current–Voltage Characteristics

The output characteristics of an NPT-IGBT, shown in Fig. 2.10, consist of two operating regions. The forward current-voltage characteristics are plotted in the first quadrant, while the reverse current-voltage characteristic is plotted in the third quadrant. The PT-IGBT characteristics are like the MOSFET curves showing saturation behavior (Fig. 2.11). I_{CE} increases only slightly with V_{CE} beyond the channel pinch-off.

The steady-state forward characteristics of the IGBT consist of a family of curves, each of which corresponds to a different gate–emitter voltage (V_{GE}). Keeping the gate–emitter voltage fixed, the collector–emitter current (I_{CE}) is measured as a function of the collector-emitter voltage (V_{CE}). The forward characteristics closely resemble MOSFET characteristics insofar as the shape of the characteristics is concerned. A striking dissimilarity is one order of

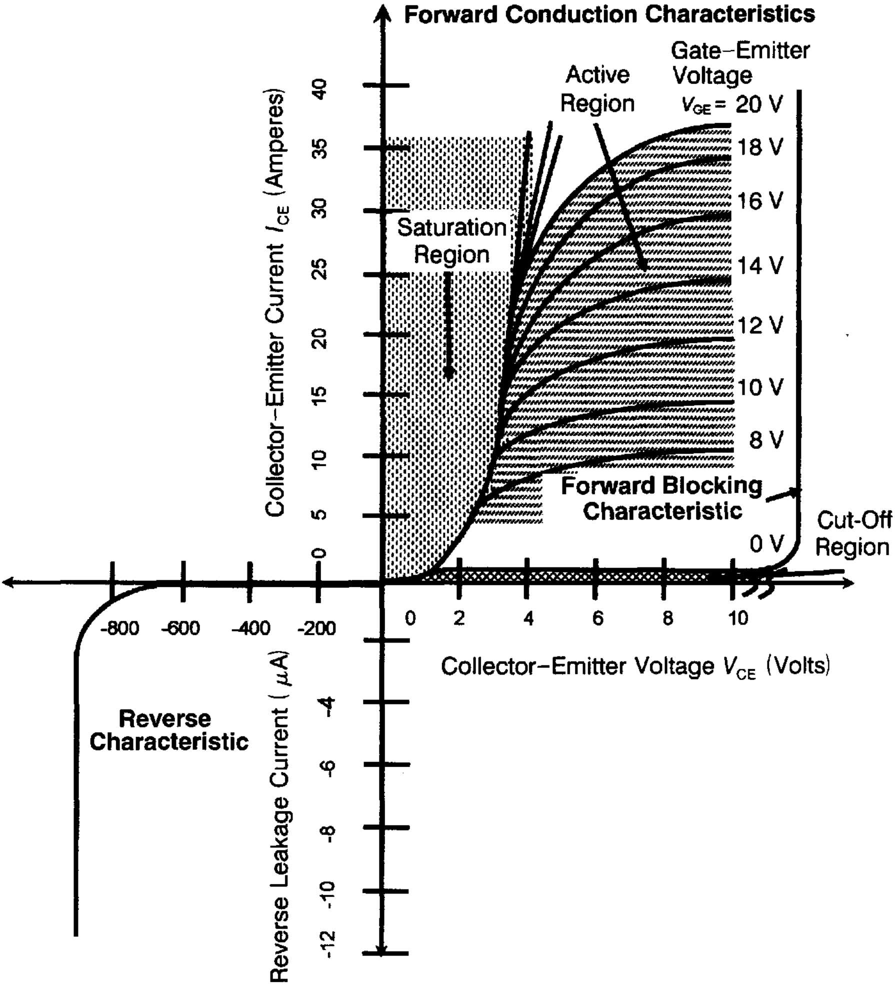


Figure 2.10 Output current–voltage characteristics of NPT-IGBT.

magnitude higher current obtained in IGBT when compared to a power MOSFET of comparable voltage and current rating. Another noteworthy distinguishing feature is the 0.7-V offset from the origin. The entire characteristic family is translated from the origin by this voltage magnitude. It may be recalled that while substituting the N^+ substrate of the MOSFET with the P^+ substrate in the IGBT, an extra P-N junction has been incorporated in the device. This P-N junction makes its operation fundamentally different from the MOSFET. The voltage drop across the IGBT is the sum of the voltage drop across the P-N junction and that across the driving MOSFET. So, unlike the power MOSFET, the voltage drop across the IGBT never falls

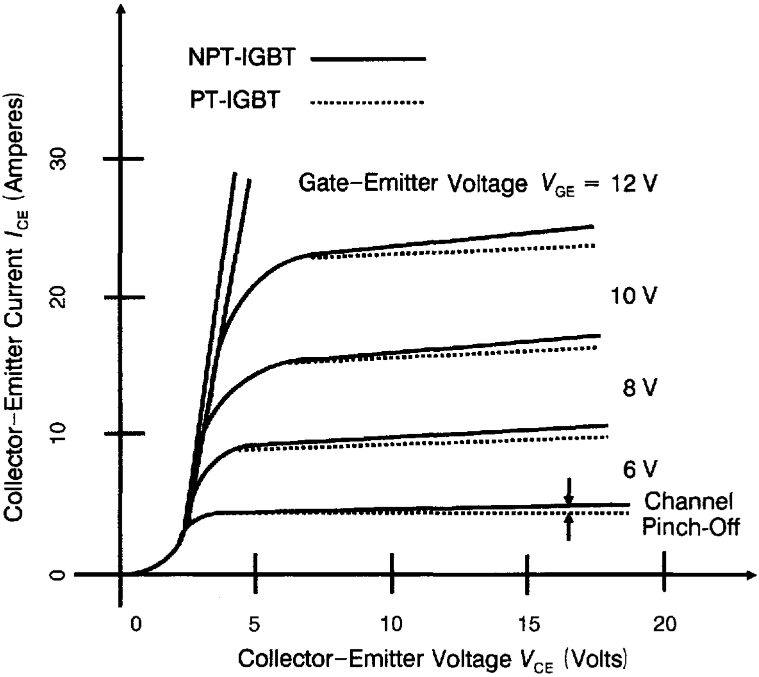


Figure 2.11 Comparison of output characteristics of NPT-IGBT and PT-IGBT.

below the diode threshold. As the voltage drop across the P^+ collector/N^--base junction of IGBT increases, the current I_{CE} rises steeply with the voltage V_{CE}, as in the linear regime of MOSFET operation. Hereafter, the channel is pinched off, and beyond this point, the current does not rise with increase of voltage V_{CE}. This is observed as a saturation of current. (The *channel pinch-off phenomenon* will be explained in connection with MOSFET operation in Chapter 3; see Section 3.3 and Figs. 3.7c and 3.7d). As the gate–emitter voltage is increased, higher saturation current levels are recorded. For a higher V_{GE} value, I_{CE} increases (Fig. 2.12) and the complete I_{CE}–V_{CE} curve is shifted upward relative to the lower V_{GE} curve. In this way, different curves are obtained for a range of V_{GE} values and the family of characteristics depicted in Fig. 2.10 is generated.

It may be remarked here that ultrafast IGBTs fabricated by heavy lifetime killing exhibit a phenomenon called "switchback," shown in Fig. 2.13, which means that the voltage drop is higher at low currents and low temperatures but decreases as either the current or temperature is raised. The term originates from the fact that when the voltage drop is measured on a curve tracer, the trace suddenly flies back to the left side of the screen as the current is raised. This behavior is observed as a "bump" in the saturation voltage portion of the current–voltage characteristics, which disappears as

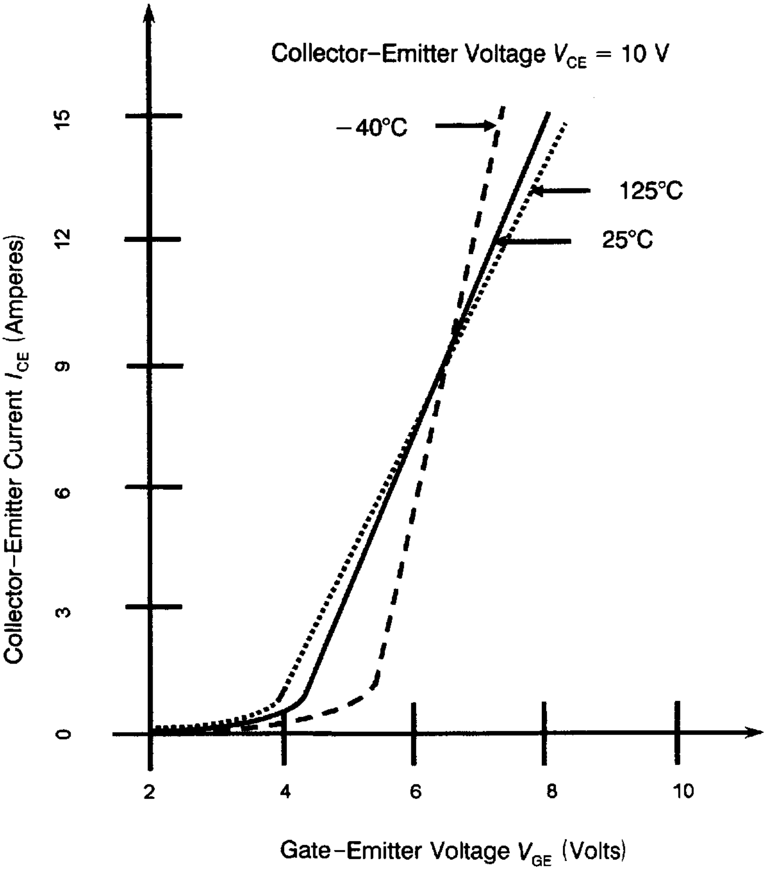


Figure 2.12 IGBT transfer characteristics at different temperatures.

the temperature is raised because the lifetime of carriers increases. The switchback is ascribed to lifetime killing, which delays the onset of conductivity modulation. The modulation effect is small at low current levels and temperatures, thereby increasing the forward drop. But it becomes significant as the current or temperature is increased, lowering the forward drop.

The reverse characteristic is identical to that of a P-N junction diode. The reverse current is extremely low (~ nA) at low voltages but increases sharply near the breakdown voltage.

2.3.2 Transfer Characteristics of IGBT

The transfer characteristics of IGBT, shown in Fig. 2.12, portray the variation of I_{CE} with V_{GE} values at three different temperatures, namely, 25°C, 125°C, and −40°C. The gradient of the transfer characteristic at a given tempera-

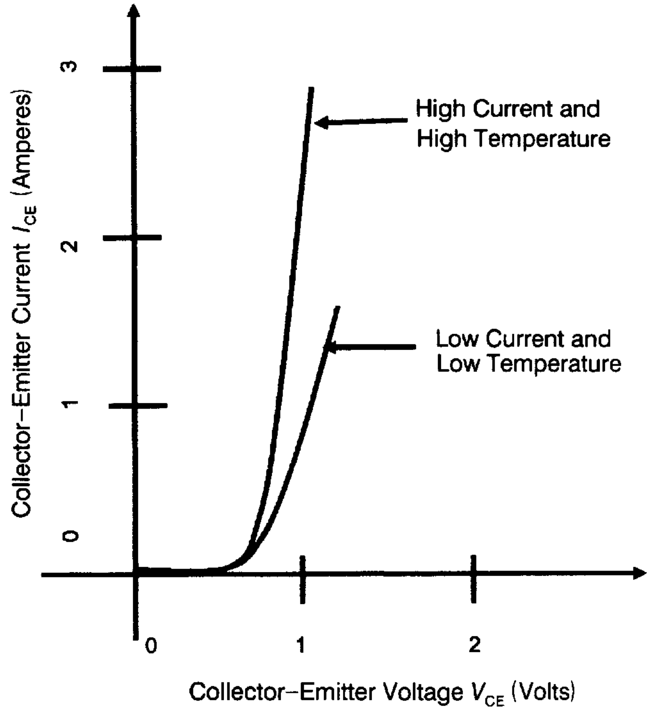


Figure 2.13 Switchback phenomenon in an IGBT with minority-carrier lifetime killing.

ture is a measure of the transconductance (g_m) of the device at that temperature:

$$g_m = \left.\frac{\partial I_{CE}}{\partial V_{CE}}\right|_{V_{CE}=\text{Constant}} \tag{2.2}$$

For a high current handling capability with low gate drive voltage, the g_m must be large. Design of the channel and gate structure dictates the g_m of IGBT. Both g_m and the ON resistance of IGBT are controlled by the channel length, determined by the difference in diffusion depths of the P base and N^+ emitter for double-diffused (DMOSFET) IGBT structure. For P-base surface concentrations $< 1 \times 10^{18}$ cm^{-3}, the depletion layer stretches more than 1 μm into the P base. Hence channel lengths smaller than 1 μm are permissible only for low voltages.

The IGBT transfer characteristic is similar to the MOSFET transfer curve except for the higher current level in the IGBT. The point of intersection of the tangent to the transfer characteristic determines the threshold voltage (V_{Th}). Alternatively, the square root of collector–emitter current ($\sqrt{I_{CE}}$) is plotted against V_{GE} for a given V_{CE}. The point of intersection of the resulting linear graph with the V_{GE} axis yields the threshold voltage. V_{Th} is also defined as the gate–emitter voltage (V_{GE}) value required to produce collector–emitter current (I_{CE}) = 1 mA at collector–emitter voltage (V_{CE}) = 10 V. Threshold voltage V_{Th} values for IGBTs are around 3–4 V. In general terms, for a given application, the gate–emitter voltage required to obtain a specified collector–emitter current at a fixed collector–emitter voltage is the threshold voltage. Both the collector–emitter current and the collector–emitter voltage are chosen according to the convenience of the customer.

It is worthwhile noting here that the current-carrying capability of a power semiconductor device is also determined by thermal or transconductance constraints. Like the current gain of a bipolar transistor, the transconductance of an IGBT increases with collector current, flattening out at a peak level for a range of collector currents and then decreases with collector current. Frequently, bipolar transistors cannot operate at the current level permitted by thermal limitations because their current gain has already fallen much below the peak value. But in the IGBT, the peak transconductance occurs at collector currents much higher than the thermally limited transconductance. The reason is that in the IGBT, the transconductance flattens out when the saturation phenomenon in the MOSFET decreases the base current drive of the PNP transistor, together with the flattening of the gain of PNP transistor itself. Upon raising the temperature, the MOSFET current reduction effect dominates over the increase in gain of PNP transistor. Hence, at higher temperatures the transconductance ceases to increase at a lower collector current.

Moreover, because lifetime killing decreases the current gain of the PNP transistor, the tranconductance of fast IGBTs peaks at a lower collector current than that of the slow ones. This, however, is a secondary effect because the major factor controlling the current gain of the PNP transistor is the N-buffer layer. Protection of the IGBT, under short-circuit conditions, is provided by the decrease in transconductance with both current and temperature. However, this reduction is very slight and may not often be good enough for protection.

2.4 SWITCHING BEHAVIOR OF IGBT

2.4.1 IGBT Turn-On

Since the input terminal of an IGBT is a MOSFET gate, the turn-on signal is applied to the MOSFET gate. The turn-on takes place by *voltage* rather than *current*, but the speed of turn-on increases with the magnitude of gate current. The gate signal current is in the nA range, while the same for current-controlled devices like BJT and thyristor is in the mA range and even higher. (This is only true for DC. The gate capacitance of any reasonably large IGBT can be ~5000 pF. Depending on the switching speed, the transient current can be much larger.) Low gate power requirement means that computer-controlled signals from integrated driver circuits can be used for this purpose. Furthermore, the voltage turn-on process of the IGBT is faster than the current turn-on mechanism of the BJT, but the IGBT turn-on is slower than the MOSFET due to the necessity of forcing both the MOSFET and BJT inside the IGBT, into conduction,. The degree of turn-on of an IGBT in the active region is measured by its transconductance.

To turn on the IGBT, the input capacitance between gate and emitter is charged to a voltage V_{GE} greater than the threshold voltage V_{Th}. Generally, an IGBT is turned on by applying a gate pulse of amplitude greater than the threshold voltage, using a collector–emitter supply with a current limiting resistor to restrict the collector–emitter current. Typically, an IGBT may be turned on by applying a rectangular gate pulse, 5 μsec wide (or greater) of sufficient amplitude $V_{\mathrm{GE}} = 15\ \mathrm{V} > V_{\mathrm{Th}}$, and a 200-V DC collector–emitter supply is used with suitable load resistance (40 Ω) to limit the collector current (5 A). Immediately at the onset of the gate pulse, the gate–emitter capacitance C_{GE} starts charging and the turn-on time is < 1 μsec.

2.4.2 IGBT Turn-On with a Resistive Load

Figure 2.14a shows a switching circuit using an IGBT with a purely resistive load R [22]. Figure 2.14b illustrates the voltage and current waveforms as a function of time. Starting at the instant of time $t = 0$, a voltage v_{GE} is applied to the gate. This voltage rises according to the values of gate–emitter capacitance C_{GE} and the series resistance R_{GE}. When the gate voltage equals the threshold voltage V_{Th}, the collector–emitter current i_{CE} begins to flow and rises rapidly according to the transconductance parameter g_m given by the transfer characteristics i_{CE} versus v_{GE}. Finally, it reaches the steady-state load value $i_{\mathrm{CE}} = I_{\mathrm{L}} \cong V_{\mathrm{s}}/R$. The *turn-on time* t_{on} is the sum of *delay time* $t_{\mathrm{d(on)}}$ (time for the current to rise from 0 to $0.1I_{\mathrm{L}}$) and *rise time* t_{ri} (time for the current to rise from $0.1I_{\mathrm{L}}$ to $0.9I_{\mathrm{L}}$). It is convenient to define t_{d} as the time interval between the application of gate drive voltage V_{GE} and its rise to the threshold value $v_{\mathrm{GE}} = V_{\mathrm{Th}}$. Then, the *current rise time* t_{ri} is defined as the interval spanning from the time $v_{\mathrm{GE}} = V_{\mathrm{Th}}$ to the time at which $v_{\mathrm{GE}} = V'_{\mathrm{GE}}$, a value that sustains the steady on-state current $I_{\mathrm{CE}} = I_{\mathrm{L}}$, where

$$V'_{\mathrm{GE}} = \frac{I_{\mathrm{CE}}}{g_m} + V_{\mathrm{Th}} \tag{2.3}$$

During the time of rise of collector–emitter current i_{CE}, the collector–emitter voltage v_{CE} falls. This is because for a resistive load we have

$$v_{\mathrm{CE}} = V_{\mathrm{s}} - i_{\mathrm{CE}} R \tag{2.4}$$

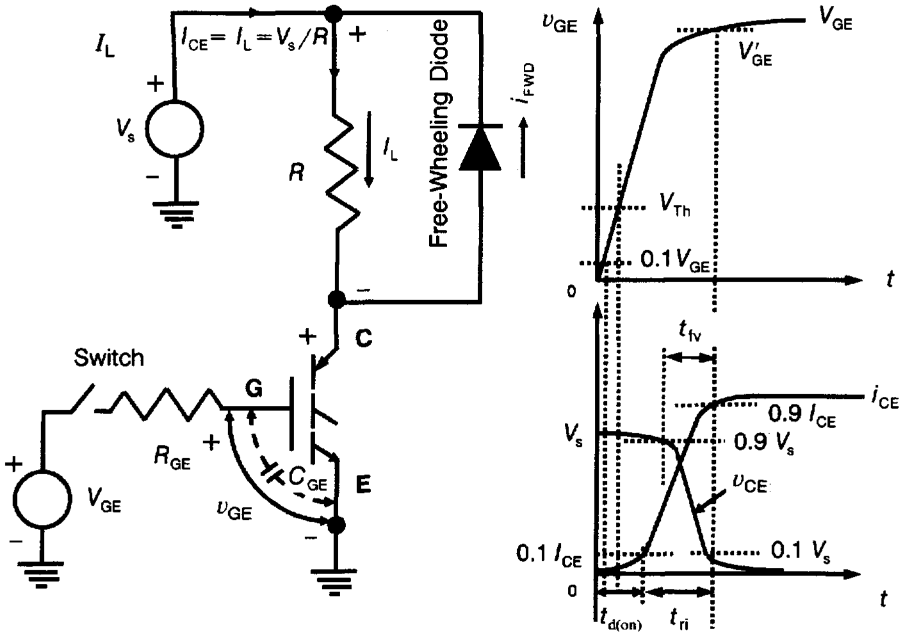


Figure 2.14 Study of IGBT turn-on with resistive load. (a) IGBT turn-on circuit with resistive load. (b) Voltage and current waveforms. **Notation:** V_s = supply voltage (V), V_{GE} = gate-supply voltage (V), C_{GE} = gate-emitter capacitance (F), R_{GE} = gate-circuit resistance (Ω), R = load resistance (Ω), i_L = instantaneous load current (A), v_{GE} = instantaneous gate-emitter voltage (V), V'_{GE} = gate–emitter sustaining voltage (V), the voltage that just maintains the steady on-state current; V_{Th} = threshold voltage (V), i_{CE} = instantaneous collector–emitter current (A), I_{CE} = steady ON-state collector–emitter current (A), t_{fv} = collector voltage fall time (sec), t_n = collector current rise time (sec), and $t_{d(on)}$ = turn-on delay time (sec).

The *voltage fall time* t_{fv} is the time interval for decrease of voltage v_{CE} from $0.9V_s$ to $0.1V_s$ value. The loss of energy during the turning on of the IGBT (E_{on}) is given by

$$E_{on} = \int v_{CE}\, i_{CE}\, dt \tag{2.5}$$

where the integration is performed over the turn-on interval. The energy dissipation during the delay time is negligibly small. At $t = t_{d(on)}$, the gate voltage reaches the value $v_{GE} = V_{Th}$ so that the collector-emitter current rise and collector-emitter voltage fall begins where $t_{ri} = t_{fv} = t_c$, the *cross-over time*. Putting $t' = t - t_{d(on)}$, the voltage and current waveforms are converted into those shown in Fig. 2.14 (c). Then we have

$$i_{CE} = \frac{I_{CE}\,t'}{t_c} = \frac{I_L\,t'}{t_c} \tag{2.6}$$

and

$$v_{\mathrm{CE}} = -\frac{V_{\mathrm{s}}t'}{t_{\mathrm{c}}} + V_{\mathrm{s}} \tag{2.7}$$

Hence the energy dissipation is

$$\begin{aligned}
E_{\mathrm{on}} &= \int_0^{t_{\mathrm{c}}} v_{\mathrm{CE}} i_{\mathrm{CE}}\, dt' = \int_0^{t_{\mathrm{c}}} \left(-\frac{V_{\mathrm{s}}t'}{t_{\mathrm{c}}} + V_{\mathrm{s}}\right)\left(\frac{I_{\mathrm{L}}t'}{t_{\mathrm{c}}}\right) dt' \\
&= \int_0^{t_{\mathrm{c}}} \left(-\frac{V_{\mathrm{s}}I_{\mathrm{L}}t'^2\, dt'}{t_{\mathrm{c}}^2} + \frac{V_{\mathrm{s}}I_{\mathrm{L}}t'\, dt'}{t_{\mathrm{c}}}\right) \\
&= \frac{V_{\mathrm{s}}I_{\mathrm{L}}}{t_{\mathrm{c}}^2}\int_0^{t_{\mathrm{c}}} t'^2\, dt' + \frac{V_{\mathrm{s}}I_{\mathrm{L}}}{t_{\mathrm{c}}}\int_0^{t_{\mathrm{c}}} t'\, dt' = -\frac{V_{\mathrm{s}}I_{\mathrm{L}}}{t_{\mathrm{c}}^2}\left[\frac{t'^3}{3}\right]_0^{t_{\mathrm{c}}} + \frac{V_{\mathrm{s}}I_{\mathrm{L}}}{t_{\mathrm{c}}}\left[\frac{t'^2}{2}\right]_0^{t_{\mathrm{c}}} \\
&= -\frac{V_{\mathrm{s}}I_{\mathrm{L}}}{t_{\mathrm{c}}^2}\left[\frac{t_{\mathrm{c}}^3}{3} - 0\right] + \frac{V_{\mathrm{s}}I_{\mathrm{L}}}{t_{\mathrm{c}}}\left[\frac{t_{\mathrm{c}}^2}{2} - 0\right] = -\frac{V_{\mathrm{s}}I_{\mathrm{L}}}{t_{\mathrm{c}}^2} \times \frac{t_{\mathrm{c}}^3}{3} + \frac{V_{\mathrm{s}}I_{\mathrm{L}}}{t_{\mathrm{c}}} \times \frac{t_{\mathrm{c}}^2}{2} \\
&= -\frac{V_{\mathrm{s}}I_{\mathrm{L}}t_{\mathrm{c}}}{3} + \frac{V_{\mathrm{s}}I_{\mathrm{L}}t_{\mathrm{c}}}{2} \\
&= \frac{V_{\mathrm{s}}I_{\mathrm{L}}t_{\mathrm{c}}}{6}
\end{aligned} \tag{2.8}$$

For IGBT operation at a frequency f, the average power dissipation can be written as

$$P_{\mathrm{on}} = E_{\mathrm{on}} \times f = \frac{V_{\mathrm{s}}I_{\mathrm{L}}t_{\mathrm{c}}f}{6} \tag{2.9}$$

Example 2.1 An IGBT (whose ratings at 125°C are: $V_{\mathrm{CE(sat)}} =$ 2.0 V, turn-on crossover time $t_{\mathrm{c}} = 30$ nsec), is used to switch power in a DC circuit using supply voltage $V_{\mathrm{s}} = 400$ V, load $R_{\mathrm{L}} = 15\ \Omega$. If the switching frequency f is 1 kHz and duty cycle m is 0.65, determine the ratio of conduction power loss and turn-on power loss at this temperature. What is the value of this ratio at $f = 1$ MHz, maintaining the same duty cycle.

Load current $I_{\mathrm{L}} = \{V_{\mathrm{s}} - V_{\mathrm{CE(sat)}}\}/R_{\mathrm{L}} = (400 - 2)/15 = 26.53$ A. If t_{ON} is the on-state conduction time, conduction power loss is given by

$$P_{\mathrm{C}} = f\int_0^{t_{\mathrm{ON}}} v_{\mathrm{CE(sat)}} I_{\mathrm{CE}}\, dt = v_{\mathrm{CE(sat)}} m I_{\mathrm{CE}} = 2.0 \times 0.65 \times 26.53 = 34.49 \text{ W} \tag{E2.1.1}$$

Power dissipated during turn-on is

$$P_{\mathrm{on}} = E_{\mathrm{on}} \times f = V_{\mathrm{s}}I_{\mathrm{L}}t_{\mathrm{c}}f/6 = 400 \times 26.53 \times 30 \times 10^{-9} \times 10^3/6 = 0.0531 \text{ W}$$

Required power loss ratio at 1 kHz is $r = P_{\mathrm{C}}/P_{\mathrm{on}} = 34.49/0.0531 = 649.53$.

At 1 MHz, taking the same duty cycle, the conduction power loss remains constant: $P_C = 34.49$ W. But $P_{on} = 53.06$ W and the ratio becomes $r = P_C/P_{on} = 34.49/53.06 = 0.65$.

2.4.3 IGBT Turn-on with an Inductive Load

Now let us consider the response of a power circuit with an inductive load (Fig. 2.15a). Suppose that the circuit time constant is $\gg$ the switching time of IGBT used. Then the load current I_L is constant. When the IGBT is off, the current is carried by the free-wheeling diode FWD [22]. Due to the conduction of FWD, the voltage across the load terminals is ~ 0 (Fig.2.15b).

The free-wheeling diode is primarily used to conduct the load current during IGBT turn-off. To understand the function of free-wheeling diode, it must be noted that in hard switching, turn-on losses increase as the recovery current rises and recovery time is prolonged. To minimize the turn-on losses and withstand surge voltages, a diode having fast, soft recovery characteristic is desirable. A diode with a snappy reverse characteristic and hence high

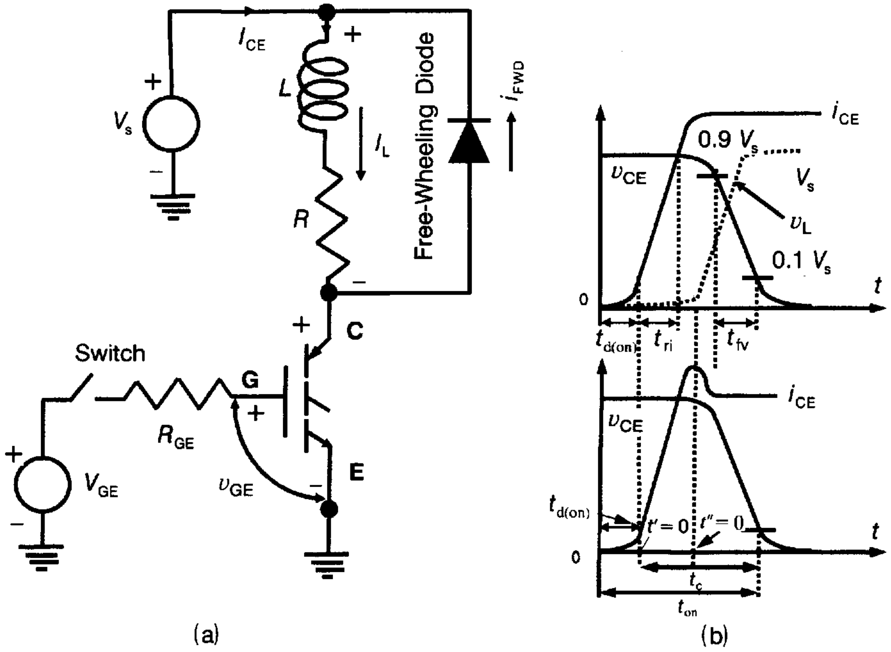


Figure 2.15 Study of IGBT turn-on with inductive load. (a) IGBT turn-on circuit with inductive load. (b) Voltage, current, and converted waveforms. **Notation:** L = inductance (H), R = resistance (Ω), t_c = crossover time (sec) from $i_{CE} = 0.05 \times$ load current to $v_{CE} = 0.05 \times$ supply voltage; t_{on} = turn-on time (sec), $t_{d(on)}$ = turn-on delay time (sec), $t' = t - t_{d(on)}$ (sec); $t'' = t - \{t_{d(on)} + t_{ri}\}$ (sec); see also Figure 2.14.

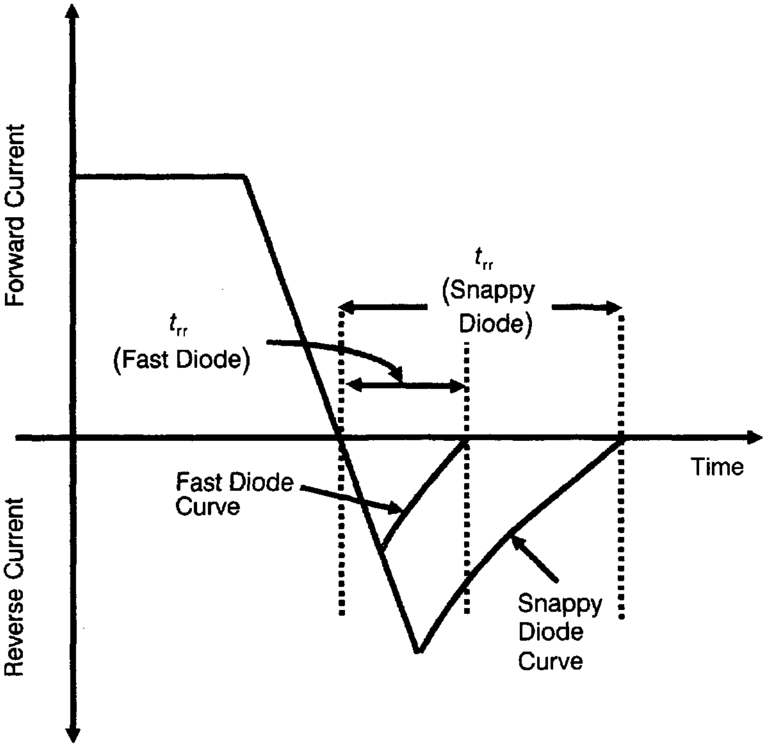


Figure 2.16 Reverse recovery waveforms of snappy and fast diodes.

recovery *di*/*dt* is problematic with transient voltages. Therefore, the turn-off current should be diverted through the free-wheeling diode instead of flowing through the IGBT as the IGBT diode may not fulfill the desired criteria. During fabrication of the free-wheeling diode, special care is taken to reduce the reverse recovery charge and reverse recovery time by decreasing the doping concentration and diffusion depth of the collector-side P layer for optimizing the ON-state carrier concentration.

Figure 2.16 compares the reverse recovery characteristics of a snappy diode with a fast recovery diode. Snappy behavior is proportional to the peak recovery current I_{RM} so that diodes with larger I_{RM} will be more snappy. Fast recovery diodes are specially tailored for both low on-state voltage drop and reduced recovery current for faster turn-off. Following are the desirable features of fast recovery diodes: (a) low forward voltage and positive temperature coefficient for safe parallel diode operation, (b) stable reverse blocking characteristics with low leakage current at high temperatures, (c) low reverse recovery losses, soft recovery, and ruggedness against dynamic avalanching, and (d) surge current capability, avalanche energy withstand capability, and a low overshoot voltage during the diode turn-on transient period.

The conventional epitaxial diode having a highly doped and relatively deep diffused emitter has several disadvantages, notably poor switching characteristics with a long reverse recovery time and a large reverse recovery charge. These diodes have also suffered from several failure mechanisms during reverse recovery due to snappy recovery and dynamic avalanching. Introduction of new lifetime killing techniques and planar junction termination designs has led to improvements in the design of fast diodes. These improvements have resulted from three important design techniques, namely; (i) emitter efficiency control, (ii) axial lifetime killing, and (iii) deep diffusion control.

The first technique is based on the use of a low P-emitter efficiency to control the gradient of the excess carrier concentrations. The shape and distribution of the stored charge in the drift region influences the reverse recovery characteristics of the diode. The increasing carrier distribution profile toward the NN^+ interface is required for achieving softer recovery characteristics and smaller reverse recovery losses compared to a flat or decreasing carrier profile. The low emitter efficiency also aids in achieving positive temperature coefficient on-state characteristics, which are necessary for parallel operation of diodes. Further control of the gradient charge profile is obtained by the use of controlled local lifetime in the drift region nearer the PN junction.

Local lifetime killing processes such as proton or helium implantation are used for controlling the axial carrier lifetime profile. This method allows a recombination layer with low lifetime near the PN junction to effectively decrease the reverse recovery parameters and produce a softer recovery without increasing the forward voltage drop. The high lifetime value near the NN^+ interface provides the additional residual charge for softer recovery. Moreover, by adding a uniform lifetime killing using electron irradiation, the softness of the diode during reverse recovery is controlled. Also, this method of carrier lifetime control does not increase the leakage current appreciably when compared to other techniques such as gold diffusion. However, the control of the gradient of the carrier concentration has still proved ineffective. Under certain combinations of forward current, commutating di/dt, circuit inductance, and junction temperature, it is likely that the diode will produce excessive voltage spikes due to snappy recovery.

Snappy recovery is normally caused by the sudden disappearance of the minority carriers stored in the drift region. This often occurs when the depletion layer reaches the N^+ region in punch-through epitaxial structures during the recovery phase, leading to current chopping, high di/dt, and hence large voltage spikes. Snappy recovery in epitaxial structure is prevented by using N-buffer layers with higher doping levels in front of the N^+ substrate. This region reduces the spreading out of the depletion layer, providing the extra charge needed for softer recovery. The doping level of the buffer layer should be sufficiently high to prevent the depletion layer from reaching the NN^+ junction but low enough for conductivity modulation.

However, these designs, in addition to the extra cost of a second epilayer, may still give snappy recovery under certain extreme conditions. By adopting a controlled deep diffused N^+ layer together with the controlled graded stored charge profile, a *progressive punchthrough* is obtained in the NN^+ interface which is practically identical to a buffer region. Thus, the effective drift region width, doping, and the punchthrough voltage are reduced safely without the diode becoming snappy. Also, by using thinner wafers, the forward voltage drop is lowered to levels near those of an epitaxial structure. This method ensures that the stored charge remains at the NN^+ interface at the latter stages of the recovery period, while the deep diffused N^+ layer prevents the depletion layer from sweeping out the remaining carriers producing soft recovery characteristics under extreme conditions. For ruggedness of performance, a suitable excess carrier profile in the drift region must be used along with maintenance of clean and uniform processes with optimum edge termination and contact designs. Lastly, oscillations taking place during reverse recovery when operating at high switching speeds must also be considered. This type of behavior is often erroneously interpreted for snappy recovery. It is due to a low level of stored charge remaining during reverse recovery. Although oscillatory recovery characteristics usually do not cause destructive voltage overshoots, they generate high electromagnetic interference (EMI), an unwanted effect in power electronics applications. The diodes thus fabricated give ultrasoft recovery characteristics with minimum EMI levels under all operating conditions.

Application of a gate signal initiates IGBT action. After the delay time t_d, the voltage v_{GE} rises to V_{Th} and the current i_{CE} increases. The diode recovers its blocking state when $i_{GE} = I_L$ and $i_{FWD} = 0$. At this instant, the voltage v_L across the load rises and that across the IGBT falls. Since practically the recovery of a diode from the conducting state to the blocking condition takes a finite time, for a short time interval after the current i_{FWD} becomes zero while the IGBT turns on, the diode remains conducting with a positive voltage applied to its emitter terminal. So the reverse current flows until the diode reverts to its blocking state. This produces a spike in the IGBT current waveform, Figure 2.15c. The higher the gate–emitter resistance R_{GE}, the slower the rise of gate–emitter voltage v_{GE} and the longer is the turn-on time of the IGBT. The associated fall time t_{fv} of the device voltage v_{CE} is affected by resistance R_{GE} and gate-to-collector capacitance C_{GC}. The rate of change of voltage is

$$\frac{dv_{CE}}{dt} = \frac{d}{dt}(v_{CG} + v_{GE}) \tag{2.10}$$

In an inductive load circuit, the voltage v_{GE} becomes nearly constant (V'_{GE}) by the time v_{CE} begins to decrease from the value V_s. Hence,

$$\frac{dv_{CE}}{dt} \cong \frac{dv_{CG}}{dt} = \frac{i_{GE}}{C_{GC}} = \frac{V_{GE} - V'_{GE}}{R_{GE}C_{GC}} \tag{2.11}$$

A high dv_{CE}/dt can increase the gate–emitter voltage to more than the threshold level causing IGBT to turn on. But the collector–emitter current flow will clamp the dv_{CE}/dt effect due to the transfer characteristic. A negative voltage applied to the IGBT during its off-state helps to nullify dv_{CE}/dt turn on.

To estimate the energy dissipation E_{on} over the cross-over time t_c of turn-on, let us neglect the effect of diode reverse recovery and put

$$t' = t - t_{d(on)} \tag{2.12}$$

and

$$t'' = t - \{t_{d(on)} + t_{ri}\} \tag{2.13}$$

Then

$$\begin{aligned}
E_{on} &= \int_0^{t'} v_{CE} i_{CE}\, dt' = \int_0^{t_{ri}} V_s i_{CE}\, dt' + \int_0^{t_{fv}} v_{CE} I_{CE}\, dt'' \\
&= V_s \int_0^{t_{ri}} \frac{I_{CE}}{t_{ri}} t'\, dt' + \int_0^{t_{fv}} \left(-\frac{V_s}{t_{fv}} t'' + V_s \right) I_{CE}\, dt'' \\
&= \frac{V_s I_{CE}}{t_{ri}} \int_0^{t_{ri}} t'\, dt' - \frac{V_s I_{CE}}{t_{fv}} \int_0^{t_{fv}} t''\, dt'' + V_s I_{CE} \int_0^{t_{fv}} dt'' \\
&= \frac{V_s I_{CE}}{t_{ri}} \left[\frac{t'^2}{2} \right]_0^{t_{ri}} - \frac{V_s I_{CE}}{t_{fv}} \left[\frac{t''^2}{2} \right]_0^{t_{fv}} + V_s I_{CE} [t'']_0^{t_{fv}} \\
&= \frac{V_s I_{CE}}{t_{ri}} \left(\frac{t_{ri}^2}{2} - 0 \right) - \frac{V_s I_{CE}}{t_{fv}} \left(\frac{t_{fv}^2}{2} - 0 \right) + V_s I_{CE} (t_{fv}) \\
&= \frac{V_s I_{CE} t_{ri}}{2} - \frac{V_s I_{CE} t_{fv}}{2} + V_s I_{CE} t_{fv} \\
&= \frac{V_s I_{CE}}{2} (t_{ri} + t_{fv}) = \frac{V_s I_{CE}}{2} t_c
\end{aligned} \tag{2.14}$$

where the energy loss during delay time has been ignored.

Example 2.2 An IGBT chopper modulates power from a 500-V DC supply into an inductive load with a free-wheeling diode connected across it. If the switching frequency is 20 kHz and the constant load current is 40 A, find the turn-on energy and power losses, neglecting the delay time loss. The ratings of IGBT are: collector current rise time = 30 nsec and collector voltage fall time = 40 nsec.

The energy loss incurred during turning on of the IGBT is

$$E_{\mathrm{on}} = \int_0^{t_{\mathrm{on}}} v_{\mathrm{CE}} i_{\mathrm{CE}}\, dt = \frac{V_{\mathrm{s}} I_{\mathrm{CE}}}{2}(t_{\mathrm{ri}} + t_{\mathrm{fv}}) = \frac{500 \times 40 \times (30 + 40) \times 10^{-9}}{2} = 7 \times 10^{-4}\ \mathrm{J} \tag{E2.2.1}$$

and the power loss is $P_{\mathrm{on}} = f \times E_{\mathrm{on}} = 20 \times 10^3 \times 7 \times 10^{-4} = 14$ W.

2.4.4 IGBT Turn-Off

During turn-off of the IGBT, the series resistor R_{GE} between gate and emitter provides a path for the input gate-to-emitter capacitance to discharge [22]. It must be noted that the resistor R_{GE} has minimum and maximum values prescribed by definite restriction guidelines. The minimum value of R_{GE} is specified as a value, which guarantees that the IGBT will not latch for any rated combination of current and voltage conditions during operation including a resistive or inductive load and the peak-allowed junction temperature. The maximum value of R_{GE} is determined by its influence on the initial fast segment of turn-off trace, as we shall see below.

As soon as the gate pulse ceases, the collector current shows a decline (Fig. 2.17a). The current flow in the MOSFET section of IGBT stops so that base current supply of PNP transistor is curtailed. The PNP transistor reverts to its blocking state and the IGBT is turned off until the gate pulse is reapplied. The turn-off time (t_{off}) is defined as the time taken by the collector current to decrease from 90% of its steady-state value to 10% of this value. The tailing off of collector current has been carefully investigated as it increases the turn-off losses and lengthens the dead time between conduction of two IGBTs in a half-bridge. It has been revealed that two separate regions can be clearly recognized in the turn-off transient of IGBT [24–28]. Accordingly, two different fall time values (t_{f1} and t_{f2}) can be associated with the two regions. The first portion of turn-off curve (fall time t_{f1}) is marked by a rapid fall of collector current with time. It is attributed to the discharge time constant of gate–emitter capacitance C_{GE} and corresponding resistance R_{GE}. Varying R_{GE} can alter the IGBT turn-off time (Fig. 2.17b) and so it is under the control of circuit designer.

The second sloppy portion of turn-off trace (fall time t_{f2}) exhibits a much slower fall of collector current with time. It is ascribed to decay of excess carriers by recombination in the N^- base of the IGBT. At the moment the IGBT was turned off, it was carrying a steady forward current. A large concentration of holes and electrons was established in the N^- base. There is no way in which these carriers can be removed except by recombination. Only after the decay of these carriers can the IGBT return to its forward blocking condition. Fall time t_{f2} is obviously governed by the carrier lifetime τ in the N^- base. It is this lifetime of minority carriers in the N^- base which

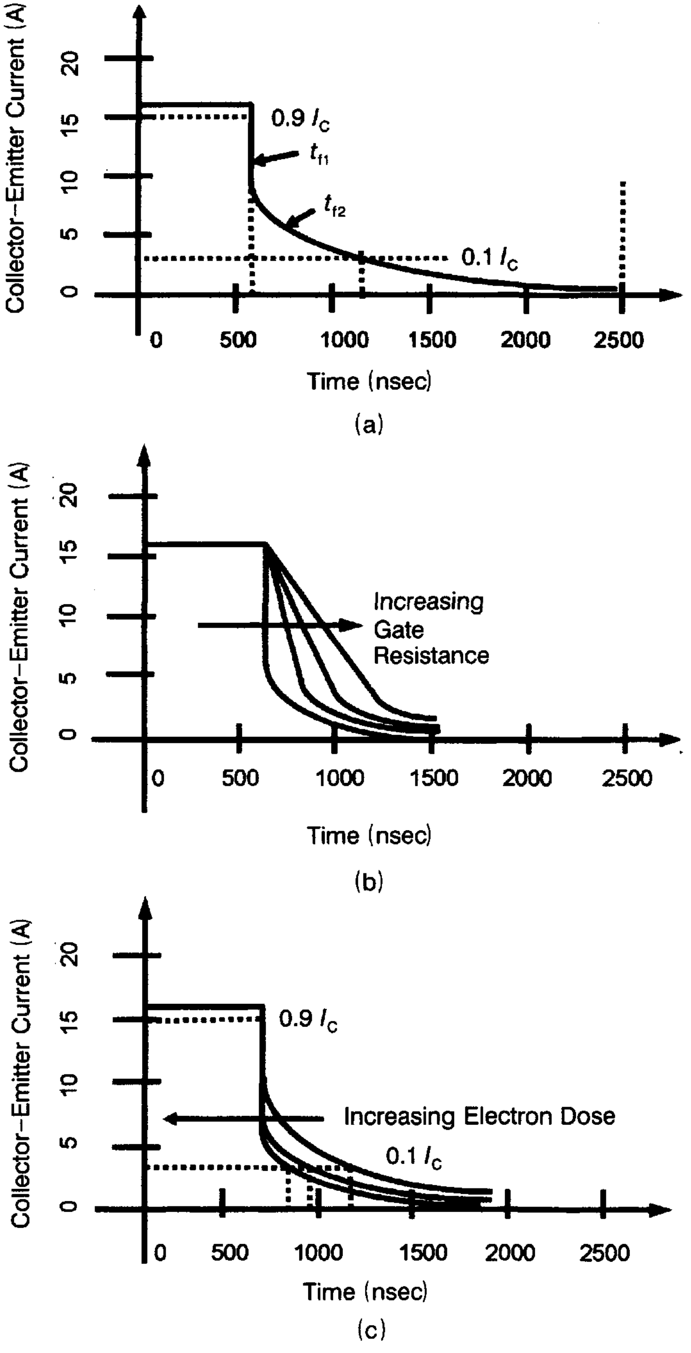


Figure 2.17 Turn-off characteristics of IGBT. (a) Typical collector current turn-off transient of an IGBT. (b) Influence of gate resistance on the turn-off characteristic of IGBT. (c) Effect of electron irradiation on turn-off characteristic of IGBT.

is a major impediment to the turn-off speed of IGBT. Since the N^- base is not accessible from outside, external drive circuits cannot be used to withdraw carriers, reducing the turn-off time. But the PNP transistor, being in a pseudo-Darlington connection, has negligible storage time. So, the turn-off time is much smaller than that for the PNP transistor in saturation mode. Nevertheless, it is insufficient for high-frequency applications. The lifetime is reduced by electron irradiation (Fig. 2.17c) or proton implantation to achieve the desired t_{f2} value; the latter provides localized lifetime killing. An N-buffer layer is also used to collect the minority carriers at turn-off speeding up the recombination rate. Both these methods decrease the gain of the PNP transistor and thereby increase the forward voltage drop of the IGBT. Heavy minority-carrier lifetime killing produces a quasi-saturation condition at turn-on, making the turn-on losses larger than the turn-off losses. Thus the turn-on losses on one hand and the latching considerations on the other restrict the gain of the PNP transistor.

Customarily, the turn-off time of IGBT is defined as the time taken by the collector–emitter current to decay from its steady ON-state value to 10% of this value. If I_{CO} denotes the steady ON-state current, the turn-off time is the time taken from the ON-state current I_{CO} to 0.1 I_{CO}. Suppose the initial sudden drop in current corresponding to the fall time t_{f1} is I_{CD}. This drop occurs due to stoppage of MOSFET channel current (I_e), and is determined by the current gain of the PNP transistor (α_{PNP}). It is written as

$$I_{CD} = I_e = (1 - \alpha_{PNP}) I_{CO} \tag{2.15}$$

Hereafter, the holes stored in the N^- base sustain the collector-emitter current. In the first stage of this process, the magnitude of this hole current equals the on-state current before turn-off. This hole current is expressed as

$$I_{CE1} = I_{CO} - I_{CD} = I_h = \alpha_{PNP} I_{CO} \tag{2.16}$$

The ensuing second stage entails exponential downfall of current at a rate decided by the carrier lifetime at high injection level, remembering that the current value is appreciable here. This current decay is written as

$$I_{CE}(t) = I_{CE1}\exp\left(-\frac{t}{\tau_{HL}}\right) = \alpha_{PNP} I_{CO}\exp\left(-\frac{t}{\tau_{HL}}\right) \tag{2.17}$$

from which the turn-off time, defined as the time taken by collector current to decrease from steady-state current to 1/10th value, is readily obtained. Thus from Eq. (2.17), we have

$$\frac{I_{CE}}{I_{CO}} = \frac{1}{10} = \alpha_{PNP}\exp\left(-\frac{t_{off}}{\tau_{HL}}\right) \quad \text{or} \quad \exp\left(\frac{t_{off}}{\tau_{HL}}\right) = 10\alpha_{PNP}$$

$$\text{or} \quad -\frac{t_{off}}{\tau_{HL}} = \ln(10\alpha_{PNP})$$

giving

$$t_{\text{off}} = \tau_{\text{HL}} \ln(10\alpha_{\text{PNP}}) \tag{2.18}$$

This derivation is based on the assumption of prevalence of high-level injection conditions in the N^- base and maintenance of a constant lifetime, irrespective of the injection level. Upon electron irradiation, both α_{PNP} and τ_{HL} are diminished, resulting in t_{off} reduction. Looking at the IGBT turn-off transient after electron irradiation, shown in Fig. 2.17c, it is clear that irradiation has a twofold effect on this waveform. Firstly, the magnitudes of the sudden drop in collector–emitter current (I_{CD}) increase. Secondly, the time span for the current tail to become zero decreases.

Example 2.3 To reduce the turn-off time of an IGBT, the electron irradiation process does carrier lifetime killing. If the pre-irradiation lifetime was 10 μsec, what will be the lifetime after subjecting the device to a dose of 1 Mrad? Radiation damage coefficient = 10^7 particles-sec/cm^2.

If the pre-irradiation lifetime is denoted by τ_{i} and the post-irradiation lifetime by τ_{f}, then τ_{i} and τ_{f} are interrelated as

$$\frac{1}{\tau_{\text{f}}} = \frac{1}{\tau_{\text{i}}} + K\phi \tag{E2.3.1}$$

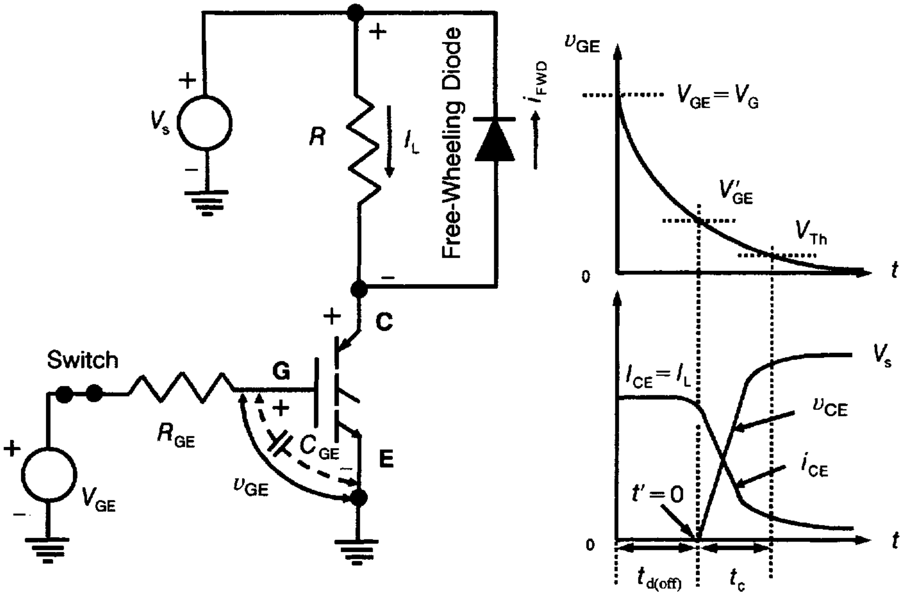
where K is the radiation damage coefficient. Now 1 Mrad equals a fluence of 1×10^{14} cm^{-2}. At these high doses, the first term is negligible and the lifetime after radiation exposure is inversely proportional to the dose. This can be expressed in terms of the radiation damage coefficient K as

$$\tau_{\text{f}} = \frac{K}{\phi} \tag{E2.3.2}$$

giving $\tau_{\text{f}} = 10^7/1 \times 10^{14} = 10^{-7}$ sec $= 0.1$ μsec.

2.4.5 IGBT Turn-Off with a Resistive Load

The turn-off mechanism of the IGBT depends on the load [22]. Usually, the IGBT turn-off is slower than that of the bipolar transistor. This is because reverse base drive applied to the BJT serves to sweep the carriers from the N^--drift region. Application of a negative gate voltage to IGBT will not help in this matter, because all recombination must occur internally. Referring to the turn-off circuit with resistive load shown in Fig. 2.18, and the simplified waveforms, turn-off is initiated by closing the gate circuit switch and reducing the gate-supply voltage V_{GE} to zero, when the gate–emitter capacitance C_{GE} begins to discharge. There is a delay time $t_{\text{d(off)}}$ before the current and voltage in the main circuit alter. The IGBT remains on until the gate voltage



(a) (b)

Figure 2.18 Study of IGBT turn-off with resistive load. (a) Resistive load turn-off circuit for IGBT. (b) Associated waveforms. **Notation:** V_s = supply voltage (V), R_{GE} = gate–emitter resistance (Ω), R = load resistance (Ω), C_{GE} = gate–emitter capacitance (F), V_{GE} = gate–supply voltage (V), v_{GE} = instantaneous gate–emitter voltage (V), V'_{GE} = gate–emitter voltage (V) maintaining the operation between saturated and active regions, V_{Th} = threshold voltage (V), $t_{d(off)}$ = turn-off delay time (sec) = time taken by the gate–emitter voltage v_{GE} to reach V'_{GE}, $t' = t - t_{d(off)}$ (sec), t_c = crossover time (sec) from the start of voltage rise to the end of current fall.

v_{GE} decreases to the value $v_{GE} = V'_{GE}$, the voltage which is just sufficient to maintain operation between saturation and active regions. As v_{GE} reduces, so does i_{CE} in accordance with the transfer characteristic, $i_{CE} = g_m(v_{GE} - V_{Th})$. Since $v_{CE} = V_s - i_{CE} R$, the voltage v_{CE} increases with fall of i_{CE}. The time t_{fi} for current fall equals the time t_{rv} for voltage rise, and these times are the same as the time t_c for crossover, reckoned from the start of the voltage rise to the end of current fall. Proceeding on similar lines as for the turn-on of the IGBT, the energy loss E_{off} during the interval t_c is calculated by putting $t' = t - t_{d(off)}$, which gives

$$E_{off} = \int_0^{t_c} v_{CE} i_{CE} \, dt' = \frac{V_s I_{CE} t_c}{6} \tag{2.19}$$

Example 2.4 In a chopper circuit, the DC supply voltage is 300 V, the resistive load is 12 Ω, and the switching frequency is 10 kHz. The IGBT used has an on-state voltage = 1.8 V, delay time = 1 μsec and crossover interval = 2 μsec. Calculate the energy loss and power loss during turn-off.

Load current is

$$I_L = \{V_s - V_{CE(sat)}\}/R_L = (300 - 1.8)/12 = 24.85 \text{ A}$$

Energy loss during delay time is

$$E_{td(off)} = V_{CE(sat)} I_{CE} t_{d(off)} = 1.8 \times 24.85 \times 1 \times 10^{-6} = 4.473 \times 10^{-5} \text{ J}$$

Energy loss during crossover interval is

$$E_{tc(off)} = V_s I_{CE} t_c/6 = (300 \times 24.85 \times 2 \times 10^{-6})/6 = 2.485 \times 10^{-3} \text{ J}$$

∴Total energy loss during turn-off is

$$E_{off} = E_{td(off)} + E_{tc(off)} = 4.473 \times 10^{-5} + 2.485 \times 10^{-3} = 2.53 \times 10^{-3} = 2.53 \text{ mJ}$$

and turn-off power loss is

$$P_{off} = E_{off} \times f = 2.53 \times 10^{-3} \times 10 \times 10^3 = 25.3 \text{ W}.$$

2.4.6 IGBT Turn-off with an Inductive Load

The circuit and the waveforms for the turn-off of an IGBT in a circuit having inductive load are illustrated in Fig. 2.19 [22]. Up to the instant the gate

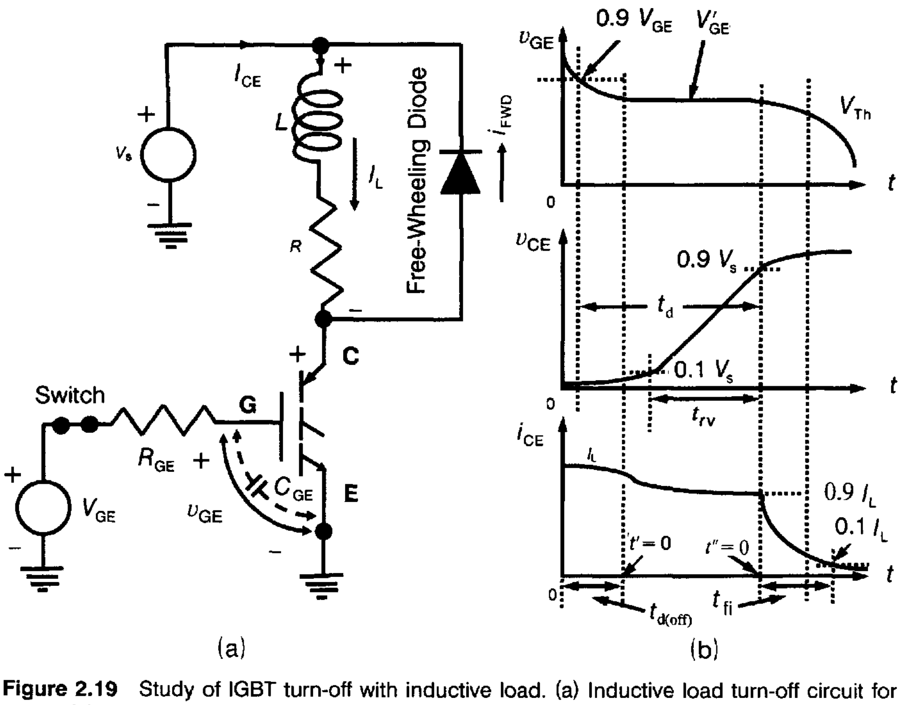

I_{CE}
+
V_s
−
+
L
I_L
R
−
Free-Wheeling Diode
i_{FWD}
+
C
Switch
G
R_{GE}
+
C_{GE}
E
+
V_{GE}
−
v_{GE}
−
v_{GE}
0.9 V_{GE}
V'_{GE}
V_{Th}
t
0
v_{CE}
0.9 V_s
t_d
0.1 V_s
t
0
t_{rv}
i_{CE}
I_L
0.9 I_L
0.1 I_L
$t'=0$
$t''=0$
t
0
$t_{d(off)}$
t_{fi}
(a)
(b)

Figure 2.19 Study of IGBT turn-off with inductive load. (a) Inductive load turn-off circuit for IGBT. (b) Current, voltage, and modified waveforms. **Notation:** L = load inductance (H), $t'' = t - \{t_{d(off)} + t_{rv}\}$. See also Fig. 2.18.

voltage falls to $\upsilon_{\mathrm{GE}} = V'_{\mathrm{GE}}$, the IGBT is conducting. Then the collector–emitter voltage υ_{CE} increases with the discharging of the capacitance C_{GE}. At the same time, the voltage υ_{GC} decreases. In this duration, the main current i_{CE} falls slightly below $I_{\mathrm{CE}} = I_{\mathrm{L}}$ in practice but ideally it should remain constant. When $\upsilon_{\mathrm{CE}} = V_{\mathrm{s}}$, the free-wheeling diode starts conducting. The load voltage υ_{L} is zero, the gate–emitter capacitance C_{GE} continues to discharge, and υ_{GE} decreases below V'_{GE}. As a result, the current i_{CE} decreases in the ratio of the transconductance. At the voltage $\upsilon_{\mathrm{GE}} < V_{\mathrm{Th}}$, the channel cannot be supported and the MOSFET is turned off. Following this abrupt fall of current i_{CE}, there is a slowly decaying tail current which is the same as the collector current in an open-base BJT circuit.

To calculate the turn-off losses during inductive load turn-off, we use the modified waveform shown in Fig. 2.19b. The delay time $t_{\mathrm{d(off)}}$ is the interval from the withdrawal of the gate-supply voltage V_{GE} to the instant that the collector–emitter voltage υ_{CE} starts rising. As the collector–emitter voltage υ_{CE} rises linearly from zero to V_{s} in time t_{rv}, the collector–emitter current is maintained at $i_{\mathrm{CE}} = I_{\mathrm{CE}} = I_{\mathrm{L}}$. Subsequently, the current decreases linearly to nearly zero in time t_{fi}. In a first-order calculation, the losses due to recombination tail current are ignored as negligibly small. Remembering that $t_{\mathrm{c}} \cong t_{\mathrm{rv}} + t_{\mathrm{fi}}$ and putting $t' = t - t_{\mathrm{d(off)}}$ and $t'' = t - \{t_{\mathrm{d(off)}} + t_{\mathrm{rv}}\}$, we have

$$E_{\mathrm{off}} \int_0^{t_{\mathrm{rv}}} \upsilon_{\mathrm{CE}} I_{\mathrm{CE}}\, dt' + \int_0^{t_{\mathrm{fi}}} V_{\mathrm{s}} i_{\mathrm{CE}}\, dt'' = I_{\mathrm{CE}} \int_0^{t_{\mathrm{rv}}} \frac{V_{\mathrm{s}}}{t_{\mathrm{rv}}} t'\, dt' + V_{\mathrm{s}} \int_0^{t_{\mathrm{fi}}} \left(-\frac{I_{\mathrm{CE}}}{t_{\mathrm{fi}}} t'' + I_{\mathrm{CE}} \right) dt''$$

$$= \frac{V_{\mathrm{s}} I_{\mathrm{CE}} (t_{\mathrm{rv}} + t_{\mathrm{fi}})}{2} = \frac{V_{\mathrm{s}} I_{\mathrm{CE}} t_{\mathrm{c}}}{2} \tag{2.20}$$

Example 2.5 In a power switching circuit, the IGBT used has voltage rise time = 0.5 μsec and current fall time = 0.3 μsec during turn-off. If the DC supply voltage is 250 V and the load current remains approximately constant at 20 A, find the energy dissipation during the crossover interval of the turn-off process.

Crossover time t_{c} = voltage rise time + current fall time = 0.5 + 0.3 = 0.8 μsec. The energy dissipated is therefore

$$E_{\mathrm{tc(off)}} = (250 \times 20 \times 0.8 \times 10^{-6})/2 = 2 \times 10^{-3} = 2\ \mathrm{mJ}.$$

2.4.7 Dependence of Turn-Off Time on Collector Voltage and Current

The relationship of turn-off time of an IGBT with collector voltage and current must be clearly understood. Interestingly enough, by raising the collector voltage, the turn-off time of IGBT increases. This is because at high collector voltages, a wider depletion layer is produced. So, more charges must be removed from the N^- base. This increases the turn-off time of the device.

On increasing the collector current, the turn-off time decreases. This is because the turn-off time is defined as the interval between 90% and 10% decay values. For larger values of collector current, the 10% decay point is lifted up and becomes more distant from the recombination tail. Because the tail accounts for the slower and longer decay span, the decay becomes faster as we are more remote from the tail, leading to a speedy recovery. It must be reiterated that this decrease of turn-off time is only an apparently illusory effect insofar as switching losses are concerned. It cannot be construed to result in smaller switching losses for the device because the tail segment of the decay waveform is mainly responsible for them, and it remains the same as before. Applying the charge-control model to an open-base transistor, current tail analysis is performed [27]. In terms of the stored base charge Q_{B}, common-emitter current gain β of PNP transistor, and base transit time τ_{B}, the base current i_{B} can be expressed as [27]

$$i_{\mathrm{B}}(t) = \frac{Q_{\mathrm{B}}}{\beta\tau_{\mathrm{B}}} + \frac{dQ_{\mathrm{B}}}{dt} = 0 \tag{2.21}$$

because i_{B} is zero for open-base turn-off. For a prototype PNP transistor operating at a current density at which high-level injection occurs in the base, the base transit time, a ratio of the base charge to the collector current can be expressed in terms of the fundamental device parameters. If γ is the emitter injection efficiency of PNP transistor, b is the ratio of electron to hole mobility $= \mu_{\mathrm{n}}/\mu_{\mathrm{p}}$, W_{B} is the thickness of N^{-}-base region, L_{a} is the ambipolar diffusion length, and τ is the minority-carrier lifetime, we can write [27]

$$\tau_{\mathrm{B}} = \frac{\{\gamma(b+1)-1\}\left\{1-\operatorname{sech}\left(\frac{W_{\mathrm{B}}}{L_{\mathrm{a}}}\right)\right\}\tau}{[2\gamma-\{\gamma(b+1)-1\}]\left\{1-\operatorname{sech}\left(\frac{W_{\mathrm{B}}}{L_{\mathrm{a}}}\right)\right\}} \tag{2.22}$$

The output collector current of the IGBT is described by the equation

$$\begin{aligned} i_{\mathrm{C}}(t) &= -\frac{Q_{\mathrm{B}}(1+\beta)}{\beta\tau_{\mathrm{B}}} - \frac{dQ_{\mathrm{B}}}{dt} = -\frac{Q_{\mathrm{B}}(1+\beta)}{\beta\tau_{\mathrm{B}}} + \frac{Q_{\mathrm{B}}}{\beta\tau_{\mathrm{B}}} \\ &= \frac{-Q_{\mathrm{B}}(1+\beta)+Q_{\mathrm{B}}}{\beta\tau_{\mathrm{B}}} = \frac{-Q_{\mathrm{B}}-\beta Q_{\mathrm{B}}+Q_{\mathrm{B}}}{\beta\tau_{\mathrm{B}}} \quad \text{or} \quad i_{\mathrm{C}}(t) = -\frac{Q_{\mathrm{B}}}{\tau_{\mathrm{B}}} \end{aligned} \tag{2.23}$$

where Eq. (2.21) has been applied. Since $\beta\tau_{\mathrm{B}}$ is generally a gradually varying parameter compared to Q_{B}, to a first approximation Eq. (2.23) can be solved for $Q_B(t) \approx Q_B(0)\exp\{-t/(\beta\tau_B)\}$. If $i_C(0)$ is the collector current at the start of the current tail—that is, the current value immediately after the initial fast drop—we have

$$i_C(t) \approx \frac{Q_B(0)}{\tau_B}\exp\left(-\frac{t}{\beta\tau_B}\right) \approx i_C(0)\exp\left(-\frac{t}{\beta\tau_B}\right) \tag{2.24}$$

This means that the collector current varies exponentially with time with a time constant τ_B dependent on the excess-carrier lifetime τ_B. For fast-switching IGBTs, $L_a < W_B$ and $\gamma \approx 1$ so that Eq. (2.22) for τ_B becomes

$$\tau_B = \frac{b\tau}{2-b} \tag{2.25}$$

Substituting for τ_B in Eq. (2.24), we get

$$i_C(t) = i_C(0)\exp\left\{-\frac{t(2-b)}{\beta b\tau}\right\} \tag{2.26}$$

which implies that the lower the lifetime value, the higher the collector current. Because less time will be consumed in falling from a given collector current to a high collector current than a smaller value, the decay rate of collector current becomes faster with decrease in lifetime. So, an inverse proportionality relationship exists between the decay rate of collector current and carrier lifetime.

2.4.8 Soft Switching Performance of NPT- and PT-IGBTs

Maximum device stresses and losses take place during the switching off of a stiff DC bus. In the soft switching concept, the DC bus is made to execute high-frequency oscillations. Hence, the bus voltage passes periodically through zero, setting up ideal switching conditions for the devices connected across the bus. The resulting topology provides smaller switching losses. Although more intricate than hard switching, it minimizes the number of power devices, reduces the size of reactive components, and increases the switching frequency. To summarize, hard switching provides simple circuit topology requiring low costs. But soft switching decreases the switching losses and increases the switching frequency. Moreover, in soft switching, the circuit complexity increases, making the systems more costly. Soft switching topology is subdivided into two classes: zero-voltage switching (ZVS) and zero-current switching (ZCS). ZVS is preferred for pulse-width modulation (PWM) and ZCS for resonant current applications.

Figure 2.20a shows the turn-on circuit of IGBT under zero-voltage switching realized by the addition of a small inductor to a conventional

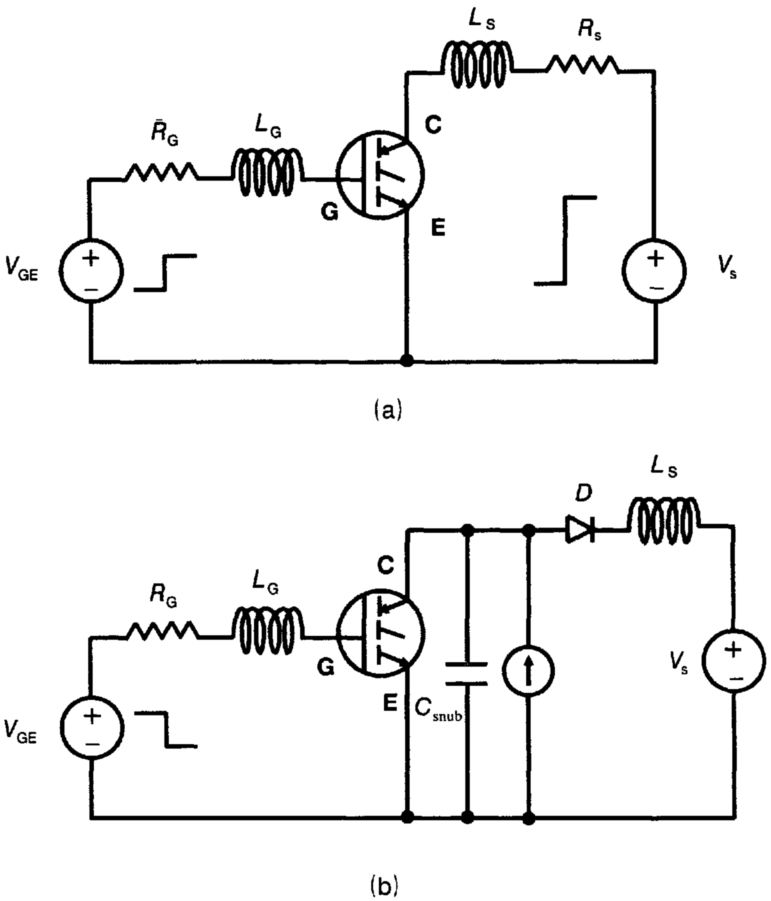


Figure 2.20 Zero voltage switching circuits of IGBT. (a) IGBT turn-on circuit. (b) IGBT turn-off circuit.

voltage source inverter circuit. Under ZVS, the IGBT is turned on when the voltage across the device is zero. First, the gate voltage V_G is applied to the IGBT and then the supply voltage is pulsed from zero to V_s. The circuit inductance L and V_s determine the rate of change of current flowing through the IGBT. Because of the application of gate voltage before V_s, the IGBT is said to turn on under zero voltage condition.

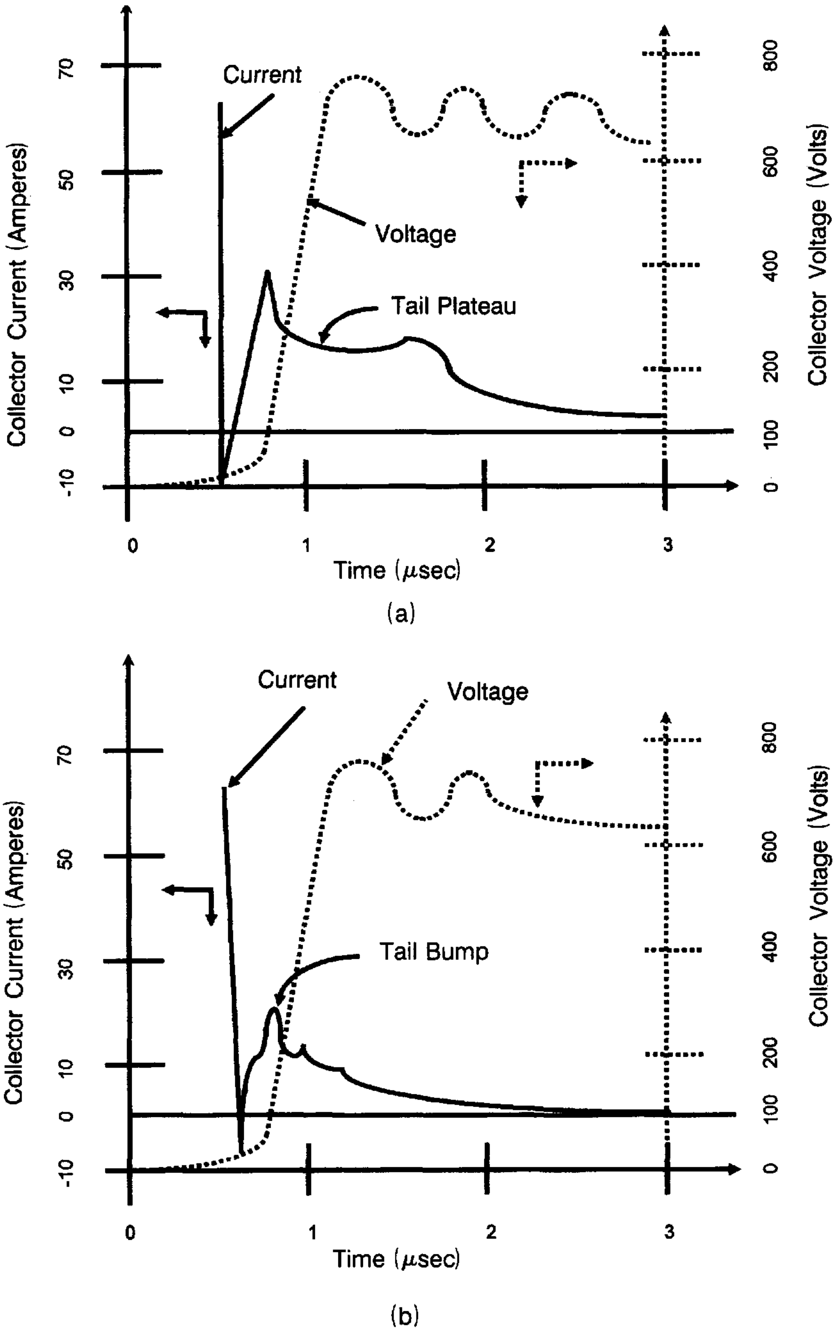
In the IGBT turn-off circuit (Fig. 2.20b), the gate voltage is withdrawn from the IGBT in the ON state. The snubber capacitor C_{snub} prevents the voltage across the IGBT from rising fast. Instead, it rises gradually to the bus voltage with the rate of rise dv/dt controlled by the C_{snub} value. As the device voltage rises slowly during turn-off, the process is said to be ZVS turn-off.

From the above discussions on IGBT turn-on and turn-off mechanisms, we must differentiate between NPT- and PT-IGBTs and analyze their behavior with reference to carrier lifetime [17, 18]. The carrier lifetime in the N^- base of an NPT-IGBT must be preserved at a high value of ~ 10 μsec during processing. This is because conductivity modulation has to take place over a large thickness of the whole wafer, and especially in the JFET region between the P-base wells where current crowding occurs due to the convergence of hole and electron currents. Despite the high lifetime, the turn-off time is not very long because the injection efficiency of P collector is low. So, back injection of electrons from the N^- base to P-collector also contributes to the turn-off process. Contrarily, in a PT-IGBT, a smaller N^--base thickness is conductivity-modulated. So, the lifetime of excess carriers need not be high. A killing process reduces the lifetime to 0.1 μsec to achieve acceptable turn-off time. It may be remarked here that in a PT device, hole injection from the P^+ collector through the N-buffer layer into the N^- base is substantially high whereas the electron injection in the reverse direction is imperceptible. So, the lifetime must be low for faster turn-off. The NPT-IGBT attains faster speed by virtue of electron back injection despite the high lifetime value.

The turn-off waveform of an NPT-IGBT shows a *tail plateau* (Fig. 2.21a). Due to the thick N^- base, the depletion region does not extend over its full length and a portion of the N^- base remains undepleted. The charges stored in this region decay by recombination, leading to a *long tail.* Contrary to the tail plateau in the turn-off transient of an NPT-IGBT, the PT device shows a *tail bump* (Fig. 2.21b). After the initial fall in current, as the voltage across the device rises to the bus voltage, the dv/dt effect supplies base current to PNP transistor. The reduction in neutral base width is accompanied by an increase in the gain of this transistor. The net result is an overall increase in carrier potential gradient, which causes an increase in hole current associated with electron current rise. These phenomena are together responsible for the peak in current. Subsequently, with the increase in voltage, the depletion capacitance decreases. Also, the electron concentration in N^- base decreases by way of recombination and their back injection into the collector. The hole concentration also falls in synchronization with the electron concentration. So, despite the narrowing of the neutral base region, the current starts falling owing to a decrease in carrier concentration.

To study the behavior of IGBT under ZCS, a single shot signal is used to gate the PT-IGBT and produce the resonant cycle of current. After the first zero crossing, the negative current is conducted by the antiparallel diode. Since the NPT-IGBT can withstand the reverse voltage, the antiparallel diode can be removed for studying these devices.

ZCS is beneficial for both PT- and NPT-IGBTs. During turning off the IGBT (PT- or NPT-), a finite amount of minority-carrier charge is stored in the N^--drift region, which causes power loss in the device. Since PT devices have a low carrier lifetime in the drift region, the stored charge recombines



(a)

(b)

Figure 2.21 Zero-voltage turn-off waveforms of IGBT. (a) Typical ZVS turn-off characteristics of NPT-IGBT. (b) Typical ZVS turn-off characteristics of PT-IGBT.

without producing a significant loss if there is sufficient interval between the turn-off and reapplication of dv/dt. Thus the constraint on the switching frequency is imposed by the recombination lifetime. The device becomes thermally unstable at high temperatures because of the increase in total stored charge.

In the NPT-IGBT, carrier lifetime is high. This IGBT has a large turn-off tail, and the stored charge is relatively insensitive to thermal variations, making the device more stable. By the time of second zero-current switching, the stored charge remains practically constant. When used with an antiparallel diode, appreciable losses occur during the rise of voltage. These losses are minimized by charge removal before voltage rise, and an efficient method of accomplishing this is by allowing a large negative current to flow through the device, sweeping away this charge. Since this device has reverse blocking capability, the antiparallel diode can be withdrawn, permitting the flow of a large negative current through the device. If an NPT-IGBT is turned off at second zero current crossing, the stored charge is swept away by the negative current. So, only a small amount of charge is left for removal by the rising voltage, thereby reducing the turn-off losses.

2.4.9 Paralleling Considerations

It is worthwhile to mention here that during parallel operation of IGBTs (Fig. 2.22), attention must be paid to both the static and dynamic considerations [28, 29]. The static problem concerns the balancing of the magnitudes of the individual collector currents. The dynamic problem relates to the choice of chips with equal turn-on and turn-off times apart from identical collector current magnitudes. For best current sharing, the IGBTs must be matched for threshold voltage and transconductance values. Also, collector and emitter areas and topological layouts must be similar. In low-frequency applications, the dynamic unbalance originating from differential switching times is insignificant. Only static current unbalance causes concern. At high frequencies, the difference of switching times has a conspicuous role to play as well as the collector current magnitudes.

For successful parallel operation of IGBTs, certain guidelines must be followed. The guidelines for paralleling discrete IGBT devices are as follows: (i) Each IGBT must have its gate resistor. (ii) The layout of IGBT chips should be arranged in such a manner that the current flow paths are symmetrical. (iii) For thermal coupling, the parallel parts must be mounted close to each other on the same heat sink. (iv) All the paralleled components of one switch group should be fabricated from the same batch of silicon wafers. (v) Threshold voltage should be matched to within +0.1-V tolerance and saturation voltage at normal operating current to within +0.05 V. (vi) The forward voltage drop of the antiparallel diode, if any, should be matched to within +0.1 V. (vii) In case of difficulty in matching in step v, a resistor = 0.2 V/nominal current per IGBT must be included in series with emitter to force current sharing.

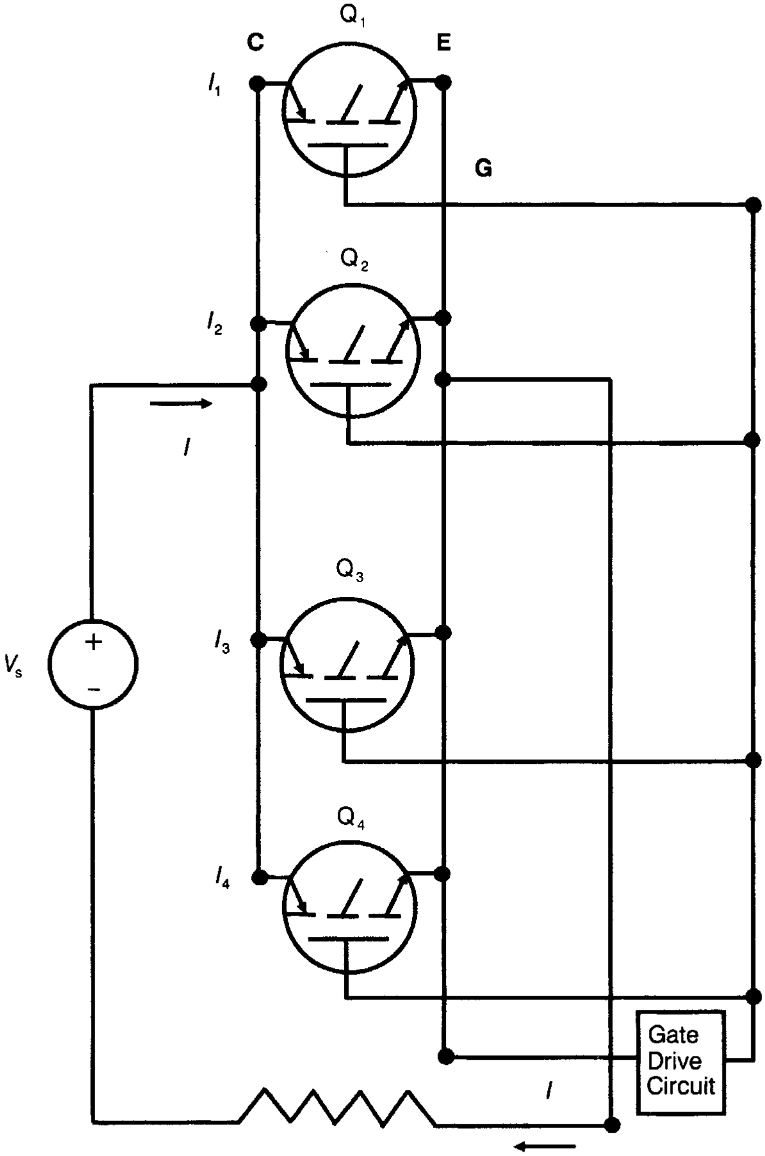

Q1
C
E
I1
G
Q2
I2
I
Q3
Vs
+
−
I3
Q4
I4
Gate Drive Circuit
I

Figure 2.22 Four IGBTs in parallel, sharing the load current.

2.5 SAFE OPERATING AREA (SOA)

The manufacturers data sheets contain the safe operating area (SOA) of IGBT [30–38]. It is defined as the area enveloped by the maximum collector–emitter voltage V_{CE} and collector-emitter current I_{CE} within which the IGBT operation must be confined to protect it from malfunction and

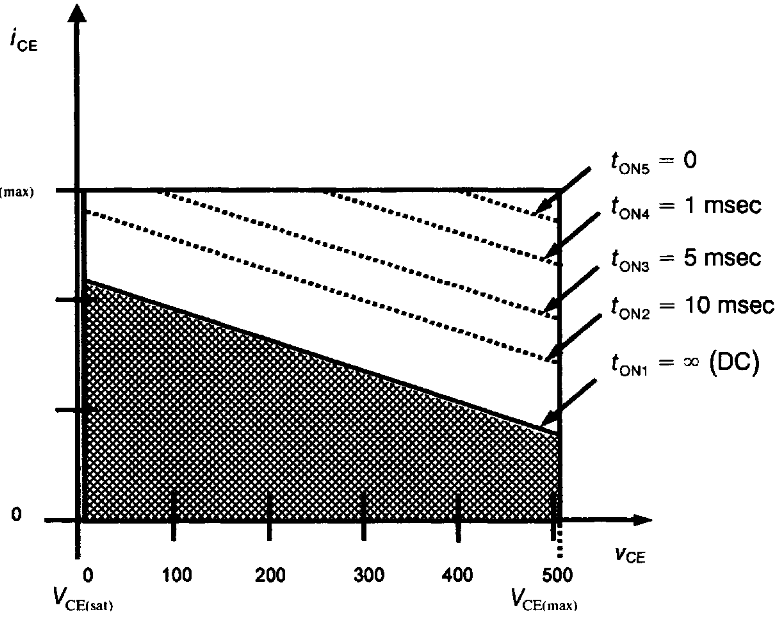


Figure 2.23 Safe operating area (SOA) of IGBT.

damage (Fig. 2.23); broken lines in the diagram depict permissible pulsed operation limits. Safe operating area gives an estimate of the peak power dissipation endurance of the device for which the junction temperature will lie below 150°C if the case temperature is 25°C. This area is mainly of interest when the IGBT is operating in the active region. For IGBT switches, the steady operating zones are the cut-off and saturation zones. The active region is traversed for a short time at turn-on and turn-off.

We shall define the three different boundaries of SOA governed by distinct destructive mechanisms: (i) Applying high V_{CE} at small I_{CE} values, the peak collector–emitter voltage withstanding capability is dictated by the edge termination employed. (ii) At high I_{CE} with low V_{CE}, the maximum collector–emitter current is determined by the latching of the four-layer PNPN parasitic thyristor in the IGBT structure. This is referred to as *current-induced latchup* because it occurs when I_{CE} becomes higher than a critical value independent of V_{CE} for low V_{CE}. (iii) When both I_{CE} and V_{CE} become simultaneously high, failure of the integral PNP transistor takes place due to thermal limitations of the chip, package, and heat sink or by the second breakdown mechanism. These thermal phenomena take place when the power dissipation is high for an extended period. For shorter-duration voltage and current stresses, second breakdown is usually the phenomenon of device failure, whereby the collector–emitter electric field exceeds the critical field ($\sim 10^5$ V/cm for Si). For avoiding second breakdown, either the resistivity of the epitaxial region is increased, or the current gain of the bipolar transistor is lowered. Thus, the open-base SOA of the integral bipolar transistor becomes larger, and second breakdown is mitigated, enhancing the SOA capability of IGBT. With regard to the second breakdown, there are two types of SOA, namely, *forward-biased safe operating area* (FBSOA) and *reverse-biased safe operating area* (RBSOA).

The maximum collector–emitter blocking voltage, at saturated I_{CE}, defines FBSOA of the IGBT. During this operational mode, mobile carriers (electrons and holes) traversing the N$^-$-base region of IGBT under the high electric field are accelerated to their respective saturated drift velocities υ_{sn} and υ_{sp} so that the free carrier densities n, p are related to the corresponding current densities J_n, J_p as

$$n = \frac{J_{\mathrm{B}}}{q\upsilon_{\mathrm{sn}}}, \qquad p = \frac{J_{\mathrm{p}}}{q\upsilon_{\mathrm{sp}}} \tag{2.27}$$

And the total positive charge in the N$^-$-base region is simply

$$Q^{+} = N_{\mathrm{D}} - \frac{J_{\mathrm{n}}}{q\upsilon_{\mathrm{sn}}} + \frac{J_{\mathrm{p}}}{q\upsilon_{\mathrm{sp}}} \tag{2.28}$$

where N_{D} is the doping density of the N$^-$ base. This charge Q^+ determines the distribution of electric field in the N$^-$ base of IGBT. The magnitude of Q^+ under FBSOA condition is much larger than in forward blocking state because $p \gg n$. Hence the electric field in the N$^-$ base increases, and the breakdown voltage of the IGBT cell is lower than that of the edge termination. By solving Poisson's equation for a positive charge Q^+ in the N$^-$ base, for a one-sided abrupt junction approximation the breakdown voltage determining the FBSOA is expressed as

$$BV_{\mathrm{SOA}} = \frac{5.34 \times 10^{13}}{\left(Q^{+}\right)^{0.75}} \tag{2.29}$$

Considering the current gain (α_{PNP}) of the open-base transistor, FBSOA limit is set by the equality

$$\alpha_{\mathrm{PNP}} M = 1 \tag{2.30}$$

where α_{PNP} and M are given by

$$\alpha_{\mathrm{PNP}} = \left\{\cosh\left(\frac{W}{L_{\mathrm{a}}}\right)\right\}^{-1}, \qquad M = \left\{1 - \left(\frac{V}{BV_{\mathrm{SOA}}}\right)^{n}\right\}^{-1} \tag{2.31}$$

W is the undepleted width of N$^-$ base and $n = 4$–6. The above equations clearly demonstrate that the avalanche breakdown of the IGBT cell is initiated at lower collector–emitter voltages with rise of collector–emitter current.

The RBSOA plays a significant role during turning off the IGBT. In this condition, the holes are the solitary charge carriers, so that the aggregate charge Q^+ contains only the additive term due to holes and the subtractive term arising from electrons is absent. In terms of the total collector–emitter current density J_{CE}, we have

$$Q^+ = N_D + \frac{J_{CE}}{qv_{sp}} \tag{2.32}$$

Consequently, the electric field distribution in the N^- base becomes worse than in FBSOA.

2.5.1 Instabilities Due to Gate Voltage Oscillations [34, 35]

IGBTs are inherently unstable and suffer from gate voltage oscillations. These oscillations are caused by *negative gate capacitance*, which appears at high collector voltages and elevated temperatures. The phenomenon has been studied [34, 35] by measurements of gate charge Q_{GE} for varying gate–emitter voltage V_{GE}, keeping the collector–emitter voltage, V_{CE}, fixed. It is found that as the gate–emitter voltage V_{GE} is increased from negative to positive values, then, at a voltage above threshold level and at a high V_{CE} value, the gate charge decreases with increasing V_{GE} rise, leading to negative capacitance values. This negative capacitance occurs because V_{GE} is less than N^--base potential, such that holes injected into the N^- base by the P^+ collector flow toward the N^- base/SiO_2 interface, building up an accumulation layer. The positive charges in the accumulation layer induce a negative charge on the gate electrode, resulting in a negative gate capacitance.

The negative gate capacitance has the same effect as an equivalent inductance. The oscillations can be modeled by introducing the delay time, τ_d, of the PNP transistor action inside the IGBT. This time delay causes changes in the AC gate capacitance, which increases with frequency from the negative value and becomes positive at a particular frequency. Both the conditions of high collector voltage and high temperature increase the current gain of the PNP transistor, thereby increasing the negative gate capacitance and inducing oscillatory disturbances. Optimization of the design of N^-base and P^+ collector increases the delay time τ_d, thereby improving the stability. Furthermore, by decreasing the gate resistance and stray inductance, the oscillations are dampened.

2.5.2 Reliability Tests [36–38]

Two commonly used benchmark tests of reliability of IGBTs under severe stress testing include:

(i) *Short-Circuit Test* Here the device carries a large current at the full bus voltage. As the drift region supports the high collector–emitter voltage, a high electric field exists in this region. Therefore the electrons and holes are rapidly accelerated to their saturation velocities resulting in a significant impact ionization rate at the junction. The ionization rate is defined as the number of electron–hole pairs generated by a carrier (electron or hole), per unit distance traversed by it. Naturally, the power developed due to the large current flow in a high electric field, causes heating of the device. The impact ionization rate is somewhat lowered by this heating. Still more carriers are liberated due to enhanced thermally induced carrier multiplication. Current crowding then takes place, underneath the gate. As the gate oxide is a poor conductor of heat, the resulting heat produced cannot be easily dissipated to the surroundings. Thus the massive thermally aided carrier multiplication, at the curved portion of the junction between P-base and N^- drift region, makes the local temperature intolerably high. Eventually, the IGBT is destroyed by heat produced due to *thermally assisted carrier multiplication at the curvature regions*. It may be remarked here that thyristor latching is a similar mechanism contributing to this type of failure. During operation in short circuit, the current flowing through the IGBT is limited by the applied gate–emitter voltage and also the transconductance of the device. It can attain values that are an order of magnitude higher than the continuous rating of the device. This high current flowing underneath the N^+ emitter contact creates a potential drop across the P-base resistance large enough to turn on the parasitic NPN transistor, resulting in latching of the IGBT. The latching is avoided by decreasing the P-base resistance and/or the device transconductance. Separate IGBT structures are therefore optimized for short-circuit or low conduction-loss operation.

(ii) *Clamped Inductive Load (CIL) Switching:* Many power electronic applications of IGBTs (e.g., power conversion, motor drives, etc.) employ inductive loads such as induction motors. For the commonly used circuit topologies, these devices have to switch on the current while the full blocking voltage is still being applied [37, 38]. For a short interval of time, $\sim 1\ \mu$sec, the IGBTs have to handle high current and high voltage simultaneously. Similar phenomena occur during turn-off. Inductive turn-off is commonly referred to as "clamped I_L." In hard-switching circuit topologies, the inductance compels the flow of a constant current in the IGBT even after turning off. The use of gate resistors to slow down the turn-off dV/dt and maintain some level of electron current avoids a potential dynamic latching condition.

The CIL turn-off process consists of two phases. In the first phase, called the *constant boundary phase*, the current flowing through the IGBT remains constant, while the voltage across it rises until it is clamped at the bus voltage. So, the device carries the full rated current while the voltage across its terminals rises to the full rated value, unless current is diverted through some alternative path. During the second phase known as the *voltage bound-*

ary condition, the voltage across the IGBT remains constant while the current flowing through it decays to zero, completing the turn-off process.

Every switching cycle is accompanied by the dissipation of a certain amount of energy by the IGBT. The same holds for the ON state when the device carries the full rated current with a forward drop of a few volts. To illustrate, during turn-on and turn-off, power levels reach up to 0.1–1 MW/cm^2. In the conducting state, a single high-power IGBT having an active area of 1 cm^2 can carry up to 60 A with a forward drop of 2–3 V so that there is a continuous power dissipation $\approx$ 100–200 W/cm^2. For comparison, a Pentium microprocessor having an active surface area of 2 cm^2 dissipates a maximum of 35 W giving a power density of 17.5 W/cm^2. As the IGBT device may not be able to forebear the heat generated by the turn-on, turn-off, and conduction processes, thermal management and thermomechanical stresses are crucial to IGBT package design.

During turn-off, the electron current falls as the MOS channel vanishes. So, the current is dominated by hole transport through the drift region. Electric charge distribution in this region is affected. As the increase in electric field at the reverse-biased junction enhances the impact generation rate, the temperature rises, showing the highest value under the emitter. The proximity of this region to the metallic contact and the package enables more efficient heat removal to the ambient. But the impact generation may overtake this heat spreading. The device fails—for example, by melting of the bonded wires. Thus, in CIL switching, the IGBT fails by *thermally assisted carrier multiplication in the parallel plane junction*, between P-base and N$^-$-drift layer. CIL switching failure represents a less severe condition than short-circuit failure because of the location of the hot spot near the contact and the package. By providing improved heat dissipation through the package, this propensity to fail is mitigated by locating the hot spot underneath the gate oxide and increasing the gate length, thereby spreading the heat over a larger area.

Example 2.6 Destructive failure of IGBT occurs due to avalanche breakdown of forward blocking junction in the presence of high current flow. Estimate the mobile charge concentration added to the depletion layer at a current density of 200 A/cm^2 and the resultant lowering of breakdown voltage if the N$^-$-base doping concentration is 5×10^{13} cm^{-3}.

At high voltages, current flows by motion of carriers through the depletion layer, which have been accelerated to their saturated drift velocity. The mobile charge concentration added to the dopant ions in the depletion layer in the form of minority carriers, by a current density J_F under high voltage, is given by

$$N = \frac{J_F}{q v_s} \tag{E2.6.1}$$

where q is the electronic charge and v_s is the saturated drift velocity $= 1 \times 10^7$ cm/sec. At $J_F = 200$ A/cm^2, $N = 200/(1.6 \times 10^{-19} \times 1 \times 10^7) = 1.25 \times 10^{14}$ cm^{-3}.

For the one-dimensional case, the breakdown voltage of the forward blocking junction between N^- base and P base in the absence of the mobile carriers is written as

$$V_B = 5.34 \times 10^{13} N_D{}^{-0.75} \tag{E2.6.2}$$

where N_D is the donor concentration in N^- base. In the presence of the additional charge, this equation is modified as

$$V_B = 5.34 \times 10^{13} (N_D + N)^{-0.75} \tag{E2.6.3}$$

Applying eqs. (E2.6.2) and (E2.6.3), the breakdown voltages without and with mobile charge are found to be 2840 V and 1109.8 V, respectively.

2.6 HIGH-TEMPERATURE OPERATION

IGBT temperature must be maintained below a critical level such that neither the forward nor the reverse leakage current becomes high enough to cause destruction [39]. During the device ON state, OFF state, turn-on, and turn-off, a significant amount of power is produced by the flowing current and must be dissipated to the surroundings to prevent the temperature from rising to a level at which device performance is unsatisfactory or ceases altogether. Heat is generally removed to the surroundings by connecting the IGBT to a cooling surface or heat sink. For most low to medium power applications ($I_C = 10\text{–}200$ A), the plastic package suffices due to low cost, light weight, and easy mounting in power circuits.

Temperature dependence of conduction characteristics, which is very significant in a power MOSFET, is minimal in the IGBT. In fact, it is adequate enough to ensure safe paralleling of IGBTs for operation at high current levels, under steady-state conditions. The total voltage drop across an IGBT is the sum of the diode and MOSFET components. The diode component has a negative temperature coefficient, the MOSFET component has a positive temperature coefficient, and the contributions of these components vary with current and temperature.

Due to the high carrier lifetime value in a NPT-IGBT [17, 18], its forward current–voltage characteristics exhibit a negative temperature coefficient of current; that is, the current falls with increasing temperature. The reason is that the thermal behavior of current is the combined result of carrier lifetime, mobility, built-in potential of diode, and contact resistance. The lifetime in an NPT-IGBT is already very high at room temperature, so its increase with temperature is small. Although the built-in potential decreases with rise of temperature due to increase in intrinsic carrier concentration n_i, the accompanying thermal processes of degradation of mobility and the increase in contact resistance are the dominating factors producing an overall negative temperature coefficient of current. This feature is advantageous for paralleling of IGBTs as it prevents thermal runaway.

For the PT-IGBT, the effect of the low lifetime value is that the temperature coefficient of current becomes positive. As the lifetime increases with temperature, more free carriers exist at a higher temperature leading to a larger value of current. This is aided by the increase in the intrinsic carrier concentration n_i, lowering the forward drop. The lifetime and intrinsic carrier concentration effects supersede the fall in mobility and rise of contact resistance to produce the positive temperature coefficient. This factor warns against the paralleling of such devices due to the risks of thermal inhomogeneities leading to destructive runaway.

In the NPT-IGBT, the high carrier lifetime is not affected appreciably by temperature and therefore the turn-off is not overly susceptible to thermal degradation. This makes the NPT-IGBT more thermally resistant during turn-off. The effect of temperature on turn-off time is more pronounced in a PT-IGBT. This is because the room temperature lifetime is low. The increase in lifetime with temperature combined with enhanced collector injection efficiency and higher n_i, makes the turn-off time longer and thereby diminishes the switching speed of the PT-IGBT. Thus, the dynamic performance of the PT-IGBT deteriorates with temperature. The positive temperature coefficient of fall time leads to excessive power dissipation. By choosing the resistor R_{GE} of the correct temperature coefficient, the above effect can be compensated to some degree. However, due to the imperceptible influence of temperature on the rise time of IGBT, such thermal compensation is not necessary during turn-on.

Example 2.7 For an IGBT used in a power switching circuit, the maximum forward voltage is 3.2 V at $V_{GE} = 15$ V. If the thermal resistance from junction to case $(R_{\theta JC}) = 2°\text{C/W}$, calculate the peak current I_{CE} in the continuous mode. Also find the peak current in pulse mode for a duty cycle of 0.6. Suppose that the heat sink is maintained at 30°C and the maximum permissible junction temperature is 130°C. The switching losses may be assumed to be negligibly smaller than the conduction losses at the device operating frequency.

The difference δT between the junction temperature (T_j) and heat sink temperature (T_S) is related to the average power dissipation P_D and thermal resistance $R_{\theta JC}$ from junction to case by the equation

$$\delta T = P_D R_{\theta JC} \tag{E2.7.1}$$

giving $P_D = \delta T / R_{\theta JC} = (130 - 30)/2 = 50$ W. This is the maximum tolerable power dissipation. In the continuous mode, $I_{CE} = P_D / V_{CE(sat)} = 50/3.2 = 15.625$ A. Thus the maximum allowed continuous current under the given thermal conditions is 15.625 A. In the pulse mode ($m = 0.6$, $t_{ON} = 0.6\ T$), the average power is

$$P_D = \frac{1}{T}\int_0^T v_{CE} i_{CE}\, dt = \frac{1}{T}\int_0^T V_{CE(sat)} I_{CE}\, dt = \frac{t_{ON}}{T} V_{CE(sat)} I_{CE} = m V_{CE(sat)} I_{CE} \tag{E2.7.2}$$

Hence the maximum current pulse, which is tolerated, is

$$I_{CE} = \frac{P_D}{mV_{CE(sat)}} \tag{E2.7.3}$$

giving $I_{CE} = 50/(0.6 \times 3.2) = 26.04$ A.

Example 2.8 A 500-V, 100-A IGBT having a forward drop of 2.5 V modulates power from a 200 V DC supply to a resistive load. If the maximum junction temperature is 150°C and the case is transiently maintained at 50°C, find the magnitude of 25-msec current pulse that the IGBT can survive. It is given that the transient thermal impedance $Z_{\theta JC(t)}$ is 0.075°C/W for a time of 25 msec.

Power dissipation for a current pulse of magnitude I_{CE} is

$$P_D = V_{CE(sat)} I_{CE} \tag{E2.8.1}$$

Also,

$$P_D = \delta T / Z_{\theta JC(t)} \tag{E2.8.2}$$

$$\therefore \quad V_{CE(sat)} I_{CE} = \delta T / Z_{\theta JC(t)} \tag{E2.8.3}$$

from which we obtain

$$I_{CE} = \delta T / \{ V_{CE(sat)} Z_{\theta JC(t)} \} \tag{E2.8.4}$$

Thus $I_{CE} = (150 - 50)/(2.5 \times 0.075) = 533.33$ A. This is the magnitude of the current pulse. It is obviously an uncontrollable current pulse because it is much higher than the current rating of IGBT (100 A).

Example 2.9 If the forward drop of an IGBT at the specified current (50 A) is 2 V, what thermal resistance from case to ambient will permit safe operation at the rated current. Thermal resistance from junction to case ($R_{\theta JC}$) is 0.9 °C/W. Junction temperature is 150°C and ambient temperature is 25°C.

Required thermal resistance from case to ambient is given by

$$R_{\theta CA} = R_{\theta JA} - R_{\theta JC} = (T_j - T_A)/P_D - R_{\theta JC} = (150 - 25)/\{ I_{CE} V_{CE(sat)} \} - 0.9$$

$$= 125/(50 \times 2) - 0.9 = 0.35\text{°C/W}$$

2.7 RADIATION EFFECTS

The effects of radiation on device characteristics play a crucial role in medical, space, and defense applications. It is necessary to know about the interaction of major types of radiation such as ionizing radiation or gamma rays, as well as neutron radiation with IGBT [40]. Gamma rays produce excess charge carriers and raise the temperature of the device. The effect of excess charge carriers is more influential. When an IGBT or MOSFET is subjected to gamma radiation, electron–hole pairs are produced in the gate oxide. Electron concentration decays by recombination, but holes are trapped by the large number of hole traps present in thermal oxides. This positive hole charge has the same effect as the surface-state charge. As an N-channel device has a positive threshold voltage, the presence of positive charge due to holes results in inversion by a smaller applied voltage—that is, a lower threshold voltage. Higher doses of gamma rays may impair device function, by rendering it conducting without application of voltage.

Neutron irradiation of silicon lowers the mobility (μ) and carrier lifetime (τ) in N^- base by inflicting lattice damages and creation of generation-recombination centers. Damage-induced deep-level compensation centers reduce the effective doping concentration (N_D). The combined effect of μ and N_D degradation is observed as a rise of resistivity (ρ) of N^- base. As the ON-state voltage drop of IGBT is determined by both ρ and τ, the ON-state voltage increases with neutron fluence. Naturally, the saturation current (at a given potential) shows a decline. However, the turn-off time of IGBT decreases. Thus, the IGBT suffers from more conduction losses, but the switching losses are smaller.

The operation of a MOSFET device is not sensitive to lifetime effects but increase in resistivity of the drain region makes the ON-resistance higher. Capacitance discharging, and not lifetime, controls the turn-off time of MOSFET. It remains unaltered by exposure to neutrons.

2.8 TRENCH-GATE IGBT AND INJECTION-ENHANCED IGBT (IEGT)

In the forward conduction state of the conventional DMOS IGBT, the N^--base region under the gate electrode, sandwiched between two neighboring P-base wells (the so-called *JFET region*), is the scene of intense current crowding due to both hole and electron currents. Holes injected by the P^+ collector and reaching the gate are pushed by the positive gate voltage. They move below the channel and cross into the base. Then traveling under the emitter, through the P^+ region, they reach the emitter contact where they are collected by the contact. Another important effect occurring under the gate is the formation of the N-type MOSFET channel at the surface of the P base through which the electron current flows downward into the N^- base to the P^+ collector. Furthermore, the P-base/N^- base junction is reverse-biased during ON state so that the electric field exists across the depletion region formed across the junction. Due to the *junction curvature effect*, electric field crowding occurs at the corners of the P-base cylindrical junction. Thus both *electric current crowding* and *electric field crowding* are simultaneously present in the aforesaid space. Electric current crowding does not allow sufficient conductivity modulation of this space. Consequently, the forward, or ON resistance, of the device increases. By enlarging the gate length, more freedom for conductivity modulation is provided, lowering the ON resistance and thereby decreasing the conduction losses. Moreover, the path length traversed by the electrons under the emitter must be shortened as the voltage drop here may exceed 0.7 V, causing latchup. By decreasing this path length, the susceptibility of IGBT to latchup is decreased.

The trench-gate IGBT (TIGBT) offers a solution to both the problems of reduced conductivity modulation between P-base wells and longer path length traversed by electrons, under the emitter. The trench gate structure is the most effective way to reduce the forward voltage drop of IGBTs. The TIGBT [41–51] is a MOS-bipolar structure in which the channel is formed on the side walls of a vertical groove. The use of a trench-gate structure allows enhanced excess carrier injection on the emitter side, between P-base wells, promoting conductivity modulation. Unlike the DMOS-IGBTs, the trench-gate IGBT contains no JFET region. So it eliminates the JFET effect and minimizes ON-state losses without compromising the turn-off performance. Trench-gate IGBTs work on the UMOSFET technology, which derives its name from the U-shaped groove, produced by reactive ion etching. The cell pitch of the UMOS trench-gate IGBT structure can be made comparatively smaller than either the DMOS or VMOS structure. Reduced cell size results in higher packing density. This increases the channel density (defined as channel width per unit active area) up to five times for the UMOS-IGBT compared to the DMOS-IGBT, resulting in superior ON-state characteristics for UMOS-IGBT. It also increases the safe operating area (SOA) by suppression of parasitic thyristor latchup. Less stringent tolerance of structural parameters is permissible in trench-gate IGBTs, due to ruggedness of the device, from the viewpoint of higher latchup immunity, together with the smaller forward potential drop of this structure, without compromising the turn-off performance. Hence, design flexibility is an added advantage of the trench gate.

Figure 2.24a shows the cross section of a trench-gate IGBT unit cell. A trench is etched by a fluorine-based, reactive ion etching (RIE) process in the gate area, as shown. The corners of the trench are rounded off to avoid electric field concentration, which can cause unwanted breakdown. The gate oxide is grown thermally over the sidewalls and the base of the trench surface, followed by filling the trench by polysilicon deposition. In the trench-gate IGBT, the MOS-channel is formed *vertically*. It can be seen that by digging the trench, the JFET region has been removed. The trench serves to transfer all the phenomena occurring near the Si surface in the confined space between the P-base wells, down below in the broader unit cell space. As the carriers wander freely with the wider unit cell space at their disposal, the ON resistance is automatically lowered. Moreover, conductivity modulation now becomes easier, because current crowding is avoided. This further reduces the ON resistance. For IGBTs having a high carrier lifetime in the

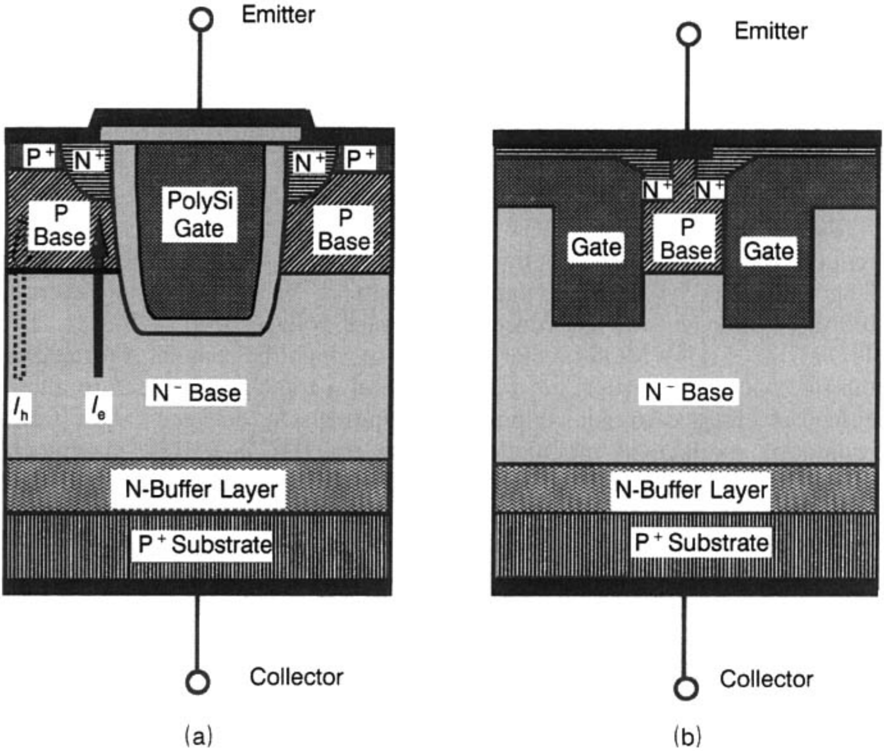

Emitter
P+
N+
PolySi Gate
N+
P+
P Base
P Base
I_h
I_e
N⁻ Base
N-Buffer Layer
P+ Substrate
Collector
(a)
Emitter
N+
N+
Gate
P Base
Gate
N⁻ Base
N-Buffer Layer
P+ Substrate
Collector
(b)

Figure 2.24 Structures of (a) trench-gate IGBT and (b) IEGT.

N^- base, the forward voltage drop at 200 A/cm^2 is 1.2 V for the UMOS structure. This is much lower than the 1.8-V drop for the DMOS IGBT. The gap of $1.8 - 1.2 = 0.6$ V between the forward drops of DMOS and UMOS structures will be further expanded by lifetime killing.

Also, it is easy to see that path of electron propagation below the N^+ emitter is now considerably shorter. This reduces the chance of the forward drop across this path exceeding the 0.7-V drop required to initiate injection across the N^+/P junction, at which undesirable thyristor action is triggered. Consequently, there is less likelihood of device latching.

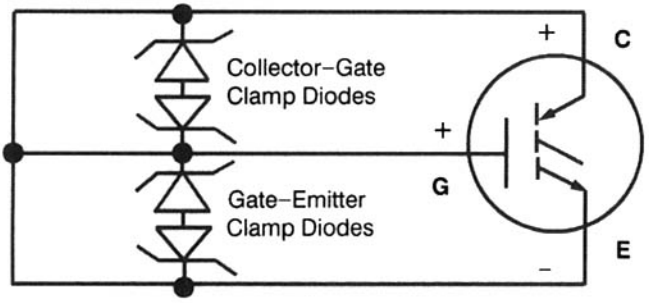
In the UMOS trench-gate IGBT, a shallow P^+ layer is sometimes diffused below the metallization short, adjacent to the N^+ emitter. This layer functions like the deep P^+ region, for latchup prevention in the DMOS IGBT structure, reducing resistance to hole current flow in the UMOS IGBT, where the resistance to hole flow is provided only by the vertical path = depth of N^+ emitter.

A useful design criterion is pursued in the injection-enhanced IGBT (IEGT) [52–55]. By *injection enhancement*, it is understood that a larger quantity of electron injection, from the N channel into the N^- base, takes place when compared to what occurs in a conventional IGBT. Based on numerical and experimental investigations, the design concept adopted in the IEGT employs deep trench (8–10 μm) gate geometry suppressing hole current. As shown in Fig. 2.24b, contacting regions of emitter electrode with P base are thinned out. Consequently, the holes have to traverse the narrow N-channel region, surrounded by the deep trench gate walls, by the diffusion mechanism for reaching the P base. Electron flow is not restricted because electrons travel in a MOS-enhanced accumulation layer along the sidewalls of the trench gate. In other words, the holes are blocked under the collector by the huge trenches, raising the potential at the drain end of the MOS channel. As a result, more electrons are injected into the drift region to recombine with the holes, producing a smaller forward voltage drop. In this manner, the IEGT achieves carrier density, like a thyristor. The device structural parameter determining the injection efficiency is W/DC, where W is the gate-to-gate distance (source width), D is the depth of the N-channel region (distance from the bottom of the trench gate to P base), and C is the cell size. Another important parameter is the electron mobility in the MOS transistor, which is controlled by the smoothness of the trench wall. Therefore, processing technology is critical to IEGT realization, which makes IEGT fabrication complicated. Another obstacle to the IEGT is the negative gate capacitance effect which causes undesirable gate voltage oscillations and instabilities as indicated in Section 2.5.1.

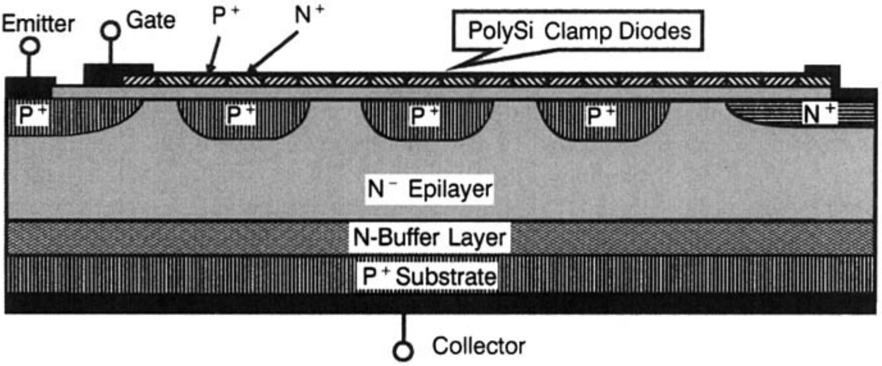
2.9 SELF-CLAMPED IGBT

A self-clamped IGBT is one that has integrated gate–collector clamp diodes [56, 57]. These IGBTs find application in automotive ignition systems where the IGBT acts as a driver. The IGBT is first turned on to ramp the current in the primary of the coil, to a predetermined value. As soon as the IGBT is turned on, the energy stored in the primary produces a voltage spike, and the secondary voltage rises to 20–40 kV until spark generation and fuel ignition take place. During a faulty condition such as an open secondary coil caused by disconnection of the spark plug, the IGBT must be protected from a high-energy impulse by clamping the gate–collector voltage below 600 V (Fig. 2.25a).

The collector–gate clamp diodes are fabricated on a polySi layer and placed across the edge termination, from the edge towards the center of the chip, in such a manner that they do not interfere with the function of the floating field rings. Figure 2.25b illustrates the concept of incorporation of the polySi diodes along one side of the field ring structure. These diodes are fabricated in a 0.5-μm-thick and 800-μm broad polySi layer, which is alternately doped P type and N^+ type. This side of the edge termination is wider than the other three sides so that the space is sufficient for the required number of diodes. To fulfill this additional space requirement, the rings are


Collector–Gate
Clamp Diodes
Gate–Emitter
Clamp Diodes
+
+
G
C
E
–

(a)


Emitter
Gate
P+
N+
PolySi Clamp Diodes
P+
P+
P+
P+
N+
N- Epilayer
N-Buffer Layer
P+ Substrate
Collector

(b)

Figure 2.25 Self-clamped IGBT. (a) Equivalent circuit of self-clamped IGBT. (b) Implementation of self-clamped IGBT in monolithic form.

made wider on this side. The performance of the field rings is not affected because the breakdown voltage achieved depends on the inter-ring spacing instead of the ring widths, beyond a certain minimum ring width. Only the ring width is increased to accommodate the diodes. In the remaining regions, the total space occupied by the field rings and the gaps between them is contracted by decreasing the ring width to that demanded by design, without changing the gaps. They thus occupy a space of only 300 μm. For minimal interaction between the clamp diodes and field rings, a close matching of the potential distributions along the Si surface and clamp diodes is essential.

2.10 RATINGS AND APPLICATIONS OF IGBT

The single IGBT device is a fully controllable, unidirectional/bidirectional semiconductor switch, which is useful in the voltage range 300–1500 V and current from 10 A to 400 A, in the frequency range from 1 kHz up to 150 kHz. By paralleling IGBTs, modules have been built with current-carrying capability of 1000 A. Breakdown voltages have been achieved up to 6500 V. Operating frequencies of up to 150 kHz, at the cost of higher forward voltage drops, have been achieved. For illustrative purposes, the specifications of a 12-A, 600-V switch mode power supply (SMPS) N-channel IGBT look like the following: continuous collector–emitter current $(I_{CE}) = 50$ A; pulsed collector–emitter current $(I_{CEM}) = 100$ A; collector-to-emitter breakdown voltage $(BV_{CES}) = 600$ V, $I_{CE} = 200$ μA, $V_{GE} = 0$ V; reverse collector–emitter breakdown voltage $(BV_{ECS}) = 10$ V, $I_{EC} = 10$ mA; collector-to-emitter saturation voltage $[V_{CE(Sat)}] = 2.0$ V, $I_{CE} = 12$ A, $V_{GE} = 15$ V; gate-to-emitter threshold voltage $V_{Th} = 5$ V, $I_{CE} = 1$ mA, $V_{CE} = V_{GE}$; gate-to-emitter leakage current $(I_{GE}) = 200$ nA, $V_{GE} = \pm 20$ V; power dissipation (P_D) at case temperature $T_c = 25°C$ is 200 W; Power dissipation derating for case temperature $T_c > 25°C$ is 1.5 W/°C; operating (T_j) and storage junction temperature $(T_{j,\,stg}) = -55$ to $+150°C$; junction-to-case thermal resistance $(R_{\theta jc}) =$ 0.7 °C/W; forward transconductance $(g_{fe}) = 5$ S, $V_{CE} = 15$ V, $I_{CE} = 5$ A; input capacitance $(C_{iee}) = 500$ pF. *Turn-on*: (i) current delay time $[t_{d(on)}] =$ 30 nsec, (ii) current rise time $(t_{ri}) = 20$ nsec, *Turn-off*: (i) Current delay time $(t_{d(off)}) = 150$ nsec, (ii) current fall time $(t_{fi}) = 100$ nsec, (iii) Turn-off energy loss $(E_{off}) = 3$ mJ.

Elaborated below are the meanings of the various terms in the IGBT data sheets.

Absolute Maximum Ratings. (i) *Continuous collector current* (I_{CE}) *at* $T_C =$ 25°C and $T_C = 100°C$ is the direct current that takes the junction to its rated temperature from the stipulated case temperature—for example, $I_{CE} = 50$ A at $T_C = 25°C$. The formula for its calculation is $I_{CE} = \Delta T/\{\theta_{j-c} \times V_{CE(ON)}@I_{CE}\}$, where ΔT is the temperature difference between the stipulated case temperature and maximum junction temperature (150°C). The ON-state voltage $V_{CE(ON}@I_{CE}$ is found with a few iterations because I_{CE} itself is unknown. (ii) *Pulsed collector current* (I_{CEM}), the peak current at which the transistor can be operated; it is well above I_C—for example, $I_{CM} = 100$ A. (iii) *Clamped inductive load current* (I_{CELM}), the maximum current which the device is able to repetitively turn off with a clamped inductive load. This rating ensures a square switching safe operating area (SOA) so that the IGBT can withstand high current and high voltage simultaneously. It is specified at 150°C and 0.8th of the voltage rating. (iv) *Maximum gate-to-emitter voltage* (V_{GE}): It is a measure of the dielectric strength and thickness of the gate dielectric. Normally, the gate dielectric quality and thickness are controlled to rupture at 80–100 V but for safe operation V_{GE} is confined up to 20–30 V to guarantee long-term reliability. (v) *Maximum power dissipation* (P_D) at 25°C and 100°C calculated by the formula $P_D = \Delta T/\theta_{j-c}$, where the symbols are as explained above in (i). (vi) *Junction temperature* (T_j): The industry standard is from $-55°C$ to $+150°C$. For details, the readers may consult the product information and data sheets of different manufacturers as mentioned in Section 1.2.

Electrical Characteristics. (i) *Collector-to-emitter breakdown voltage* (BV_{CES}) represents the lower limit of breakdown voltage, and it is defined with reference to the leakage current. The breakdown voltage has a positive temperature coefficient of 0.63 V/°C. As an example, $BV_{CES} = 600$ V at 25°C, $I_{CE} = 200$ μA, $V_{GE} = 0$ V. At 150°C, $BV_{CES} = 600 \text{ V} + 0.63 \times 150 = 694.5$ V. (ii) *Emitter-to-collector breakdown voltage* (BV_{ECS}) is the reverse breakdown voltage of the collector-base junction of PNP transistor. $BV_{ECS} = 10$ V, $I_{EC} = 10$ mA. When an IGBT is turned off, the load current flows through the diode connected in parallel across the complementary IGBT. The turn-off di/dt in the stray inductance in series with this diode produces a reverse voltage spike across the IGBT that can initiate avalanche in the junction. The reverse voltage generated is generally less than 10 V, but serious voltage spikes can result from high di/dt. (iii) *Collector-to-emitter saturation voltage* [$V_{CE(ON)}$ or $V_{CE(Sat)}$] is the key parameter to estimate the conduction losses of the IGBT. $V_{CE(Sat)} = 2.0$ V, $I_{CE} = 12$ A, $V_{GE} = 15$ V. (iv) *Gate threshold voltage* (V_{Th}) is the gate–emitter voltage at which a particular collector–emitter current starts to flow. It has a temperature coefficient of $\Delta V_{Th}/\Delta T = -11$ mV/°C. $V_{Th} = 5$ V at 25°C, $I_{CE} = 1$ mA, $V_{CE} = V_{GE}$. At 150°C, $V_{Th} = 5 - 11 \times 10^{-3} \times 150 = 5 - 1.65 = 3.35$ V. (v) *Gate-to-emitter leakage current* (I_{GE}) = 200 nA, $V_{GE} = \pm 20$ V. (vi) *Zero-gate voltage collector-current* (I_{CES}) is the limiting value of leakage current at the rated voltage. (vii) *Forward transconductance* (g_{FE}), which, unlike the bipolar transistor, increases with current but is limited thermally. It is measured by superimposing a small variation on the gate bias which takes the IGBT to the 100°C rated current in the linear mode of operation. $g_{FE} = 5$ S, $V_{CE} = 15$ V, $I_{CE} = 5$ A. (viii) *Gate charge parameters* (Q_G, Q_{GE}, Q_{GC}) which are used to size the gate drive circuit and calculate the gate drive losses. (ix) *Switching times*: *Turn-on delay time* from 10% of gate voltage to 10% of collector current, *rise time* from 10% to 90% collector current, *turn-off delay time* from 90% of gate voltage to 90% of collector current, and *fall time* from 90% to 10% collector current. For a co-pak containing the IGBT and the free-wheeling diode, the turn-off delay time is reckoned from 90% of gate voltage to 10% of collector voltage. Switching times serve as a useful guide to set up the proper dead time between the turn-off and the ensuing turn-on of the complementary IGBTs in the half-bridge circuit configuration, and the minimum and maximum pulse widths. (x) *Switching energies*: E_{ON} is the switching energy from 5% of the test current to 5% of the test voltage. E_{OFF} is measured over a period of time commencing from 5% of the test voltage and lasting up to 5 μsec. (xi) *Internal emitter inductance* (L_E) is the inductance due to the package between the bonding pad on the IGBT chip and the electrical connection at the lead. It decelerates the IGBT turn-on by a quantity varying as collector dI/dt similar to its slowing down by the Miller capacitance by an amount proportional to the collector dV/dt. For a $dI/dt = 1000$ A/μsec, the voltage produced across L_E exceeds 7 V. (xii) *Device capacitances*: *Input capacitance* (C_{iee}), the sum of gate-to-emitter and Miller capacitances; it exhibits similar voltage variation as the Miller capacitance but in a weakened form because the gate-to-emitter capacitance is much larger and voltage dependent. Other capacitances include the *output capacitance* (C_{oee}) whose voltage variation resembles that of a P-N junction diode and *reverse transfer* (Miller) capacitance (C_{ree}), which decreases with voltage but shows a complex dependence on voltage. (xiii) *Short-circuit withstand time*: The guaranteed minimum time that the device will withstand the short-circuit condition. This rating is only specified for short-circuit rated IGBTs. Here, the gate resistor cannot be smaller than the specified value. Also, the overvoltage during turn-off must be maintained at the mentioned value by clamping.

The data sheet also contains device performance curves. These curves show the output and transfer characteristics. Other data include the on-state voltage, forward transconductance, and turn-off safe operating area. Besides these, it provides derating of I_{CEM} with turn-off dV_{CE}/dt, gate-charge waveforms, switching characteristics of turn-on delay time and rise time, with respect to collector-to-emitter current, turn-off delay and fall time variation with collector-to-emitter current, and also turn-on and turn-off energy loss curves.

Surveying the product literature of different vendors, one may select the appropriate IGBT for the intended application. For power switching, IGBTs work like other devices—for example, adjusting the triggering angle α to modulate the AC voltage v_s to an AC or DC load voltage v_L. Figure 2.26

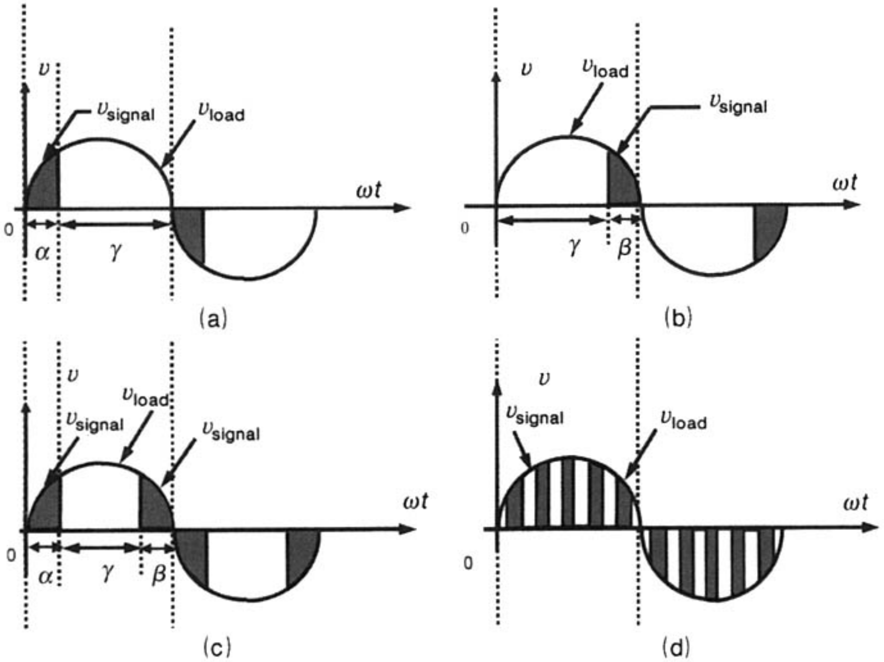


Figure 2.26 IGBT switching patterns of an AC signal. (a) Trigger angle delay = α, conduction angle = γ. (b) Conduction angle = γ, trigger angle advance = β. (c) Trigger angle delay = α, conduction angle = γ, trigger angle advance = β. (d) Pulse-width modulation.

shows the switching patterns of an AC waveform. Figure 2.26a illustrates trigger angle delay α. Figure 2.26b depicts trigger angle advance β. In Figure 2.26c, both α and β control is used with $\alpha = \beta$. Pulse width modulation (PWM) is shown in Fig. 2.26d. In all these cases, power modulation is implemented by controlling the duration of the cycle over which IGBT remains on (t_{ON}) and for which it is off (t_{OFF}). The average voltage $V_{\text{L(av)}}$ across the load is written as

$$V_{\text{L(av)}} = V_{\text{s}} \frac{t_{\text{ON}}}{t_{\text{ON}} + t_{\text{OFF}}} \tag{2.33}$$

In terms of the switching frequency f and the duty cycle $m = t_{\text{ON}}/(t_{\text{ON}} + t_{\text{OFF}})$, we may write

$$V_{\text{L(av)}} = V_{\text{s}} t_{\text{ON}} f = m V_{\text{s}} \tag{2.34}$$

For carrying out any switching function, the IGBT is toggled between two extreme states. In the ON state it is driven into saturation to attain a low voltage drop. This is because the load current i_{L} depends on the current in the N^- base of IGBT and therefore the base resistance R_{b}. A high base resistance R_{b} results in a low collector–emitter current i_{CE} and a larger on-state voltage drop, limiting the load current. By raising the gate drive $v_{\text{GE}} \sim 15$ V, R_{b} decreases due to increase in conductivity modulation of base. The BJT is thereby saturated, and only the load limits the current i_{CE}. In the OFF state, the IGBT supports the full voltage V_{s}.

2.11 SUMMARY AND TRENDS

The major families of IGBTs were reviewed in terms of their physical structure, profiles and carrier lifetimes, and the fabrication technology. The impact of these structural parameters on device performance characteristics was discussed. Both static and dynamic behavior of IGBTs were examined. It was shown that the PT-IGBTs are superior regarding the switching performance, but the same is not true from the ruggedness and thermal derating viewpoints. So, the IGBT for a given application must be chosen according to the specific needs. Any single IGBT structure is not universally optimal. Furthermore, short-circuit and clamped inductive switching tests are not equivalent for evaluating the safe operating area. After perusal of this chapter, the reader will grasp the device operational modes and their relation to the constructional details. In the succeeding chapters, the MOS and bipolar components of IGBT will be treated comprehensively to understand how the relevant terminal characteristics of IGBT are controlled.

REVIEW EXERCISES

2.1 For what type of applications are lateral IGBTs useful? Draw and explain the schematic cross section of a lateral IGBT. Point out one disadvantage of the lateral approach. In what kind of applications is the vertical IGBT structure preferred?

2.2 Differentiate between nonpunchthrough-IGBT and punchthrough-IGBT from the point of view of fabrication techniques.

2.3 Describe the various operating modes of IGBT with the help of diagrams showing the applied biasing voltages and their polarities.

2.4 Draw and explain the following for the IGBT: (a) static I–V characteristic, (b) transfer characteristic, and (c) dynamic characteristic.

2.5 Distinguish between the performance of an NPT-IGBT and PT-IGBT with reference to (a) carrier lifetime in N^- base and (b) turn-off times.

2.6 During forward blocking, which IGBT structure has a triangular electric field distribution and which one shows a rectangular shape? How does this difference influence the choice of drift region thickness for the required voltage?

2.7 (a) Explain how the decrease in forward drop of PIN diode in IGBT compensates the increase in MOSFET channel resistance accounting for the moderate variation of forward conduction characteristics of slow-switching IGBTs with temperature. Name the type of IGBT which is more thermally sensitive and why? (b) What is the advantage of small positive temperature coefficient of forward drop of IGBT at high collector currents in paralleling IGBTs?

2.8 Explain the role of the N-buffer layer in a PT-IGBT. What device parameters does it influence?

2.9 Name the IGBT variety which has a higher collector output resistance? In which class does the collector current show saturation like the MOSFET, and why?

2.10 Explain why and how the reverse blocking capability of an IGBT is sacrificed to improve the forward drop and switching speed.

2.11 Do you agree with the statement that the increase in switching speed of an IGBT must compromise with forward characteristic degradation? Justify your answer giving arguments.

2.12 Among the NPT-IGBT and PT-IGBT, which one is more vulnerable to electrical breakdown and why? In which type carrier transport takes place by drift mechanism and in which kind it is diffusion-controlled?

2.13 What are the reasons for the occurrence of two segments in the turn-off trace of IGBT? Which segment can be controlled by the circuit designer and how? Which segment can only be changed during device fabrication? What is the influence of electron irradiation on the two segments?

2.14 Why will happen if the carrier lifetime is not maintained at a high value in an NPT-IGBT? How is an NPT-IGBT able to yield a small turn off time in spite of high lifetime value?

2.15 Why must lifetime killing be used in a PT-IGBT whereas the same is not necessary in an NPT-IGBT?

2.16 Do you agree with the statement, "Carrier transport in both N- and P-channel IGBTs takes place by ambipolar diffusion and drift. So the forward drops of these devices are nearly identical." Is this statement applicable to N- and P-channel MOSFETs?

2.17 Are the forward voltage drops of slow-speed N- and P-channel IGBTs identical? Do fast-switching IGBTs of two polarities differ in their forward voltages?

2.18 Comment on the remark, "IGBT is well-suited for scaling up the blocking voltage capability."

2.19 What is meant by safe operating area of IGBT? What do the broken lines in the SOA graph represent? Define the three distinct boundaries for SOA.

2.20 Explain the phenomenon of avalanche-induced second breakdown. Define FBSOA and RBSOA for IGBT. How does the total charge in the drift region affect the electric field under FBSOA and RBSOA conditions? What is the effect of increase in drift-region charge on the breakdown voltage of IGBT?

2.21 Which reliability test, short-circuit test or clamped inductive load switching test, represents a more severe stress on the IGBT and why?

2.22 What is the effect of collector voltage and current on the turn-off time of an IGBT? Draw the forced-gate turn-off waveforms for two collector current values I_{C1} and I_{C2} ($I_{C2} > I_{C1}$). Will the switching losses due to the current tails be the same or different in the two cases?

2.23 Draw the circuit diagram for turning on an IGBT with an inductive load. Show the voltage and current waveforms. Derive an expression for the energy loss in terms of the supply voltage, collector current, and crossover time.

2.24 Illustrate IGBT turn-off with a resistive load, giving the circuit diagram, along with voltage and current variation with time. Write the equation for power loss and explain the symbols. Elaborate its derivation.

2.25 Explain the effects of gamma radiation and neutron flux on IGBT characteristics? Which of these effects is permanent and cannot be annealed out? Why?

2.26 Elucidate the *current crowding phenomenon* in the JFET region under the gate between the P-base wells in the conventional IGBT structure. How does the trench IGBT structure overcome this problem?

2.27 How does the trench-gate structure improve the channel density and forward conduction characteristics of IGBT? What are the effects of UMOS structure on the hole current path and the latching current density of IGBT?

2.28 Explain the self-clamped IGBT structure. For what applications is it needed? How are the gate–collector clamp diodes built into the IGBT without affecting the performance of the edge termination.

2.29 Give the electrical equivalent circuit and cross-sectional diagram of self-clamped IGBT showing the clamp diodes fabricated in polySi layer.

2.30 The saturation voltage and turn-on crossover time of an IGBT at 125°C are $V_{\mathrm{CE(sat)}} = 1.9$ V and $t_c = 25$ nsec, respectively. This IGBT is used to control power in a DC circuit in which the supply voltage is $V_s = 380$ V, and the load resistance R_L is 12 Ω Calculate the ratio of conductance power loss to turn-on power loss at this temperature, for a switching frequency $f = 1.2$ kHz and duty cycle $m = 0.55$. What will be the value of this ratio at $f = 1.5$ MHz, if the duty cycle remains the same.

2.31 The IGBT used in a power conversion circuit has voltage rise time of 0.3 μsec and current fall time of 0.2 μsec during turn-off. If the DC supply voltage is 180 V and the load current remains constant at 17 A, find the energy dissipation during the crossover interval of the turn-off process.

2.32 The maximum forward voltage for an IGBT used in a power conversion circuit is 3.0 V at $V_{\mathrm{GE}} = 12$ V. The thermal resistance from junction to case ($R_{\theta \mathrm{JC}}$) is 1.5°C/W. For a duty cycle of 0.5, determine the peak collector–emitter current in the continuous mode and in pulse mode. The heat sink is kept at 25°C and the maximum allowed junction temperature is 125°C. Also, the switching losses are much less than the conduction losses at the operating frequency.

REFERENCES

1. M. R. Simpson, P. A. Gough, F. I. Hshieh and V. Rumennik, Analysis of the Lateral Insulated Gate Transistor, *International Electron Devices Meeting, IEDM Technical Digest*, 1985, pp. 740–743.
2. T. P. Chow, D. N. Pattanayak, B. J. Baliga and M. S. Adler, Latching in Lateral Insulated Gate Transistors, *International Electron Devices Meeting, IEDM Technical Digest*, IEEE, New York, 1987, pp. 774–777.
3. A. Vellvehi, P. Godignon, D. Flores, J. Fernandez, S. Hidalgo, J. Rebollo and J. Millan, A New Lateral IGBT for High Temperature Operation, *Solid State Electronics*, Vol. 41, No. 5, May 1997, pp. 739–747.
4. T. P. Chow, B. J. Baliga, H. R. Chang, P. V. Gray, W. Hennessy and C. E. Logan, Vertical Insulated Gate Bipolar Transistors with Collector Short, *International Electron Devices Meeting, IEDM Technical Digest*, 1987, pp. 670–673.
5. T. P. Chow, B. J. Baliga, D. N. Pattanayak, and M. S. Adler, Comparison of P-channel LIGBTs With and Without Collector Shorts, *IEEE Electron Device Letters*, IEEE, Vol. EDL-11, No. 5, May 1990, pp. 184–186.
6. M. R. Simpson, Analysis of Negative Differential Resistance in the I–V Characteristics of Shorted Anode LIGBTs, *IEEE Trans. Electron Devices*, Vol. ED-38, No. 7, July 1991, pp. 1633–1640.
7. J. Sin and S. Mukherjee, Analysis and Characterization of the Segmented Anode LIGBT, *IEEE Trans. Electron Devices*, Vol. ED-40, 1993, p. 1300.
8. M. Gärtner, D. Vietzke, D. Reznik, M. Stoisiek, K. -G. Oppermann, and W. Gerlach, Bistability and Hysteresis in the Characteristics of Segmented Anode Lateral IGBTs, *IEEE Trans. Electron Devices*, Vol. ED-45, No. 7, July 1998, pp. 1575–1579.
9. S. Byeon, B. H. Lee, D. Y. Kim, M. K. Han, Y. I. Choi, and C. M. Yun, A Lateral Insulated Gate Bipolar Transistor Employing the Self-Implanted Sidewall Implanted Source, *International Electron Devices Meeting, IEDM Technical Digest*, IEEE, New York, 1998, pp. 687–690.

10. H. Yilmaz, J. L. Benjamin, R. F. Dyer, L. S. Chen, W. R. Van Dell and G. C. Pifer, Comparison of the Punchthrough and Nonpunchthrough Insulated Gate Transistor Structures, in *Conference Record of the IEEE Industry Applications Society (IEEE-IAS) Annual Meeting*, IEEE, New York, 1985, pp. 905–908.
11. B. J. Baliga, Analysis of the Output Conductance of Insulated Gate Transistors, *IEEE Electron Device Letters*, Vol. EDL-7, No.12, December 1986, pp. 686–688.
12. A. R. Hefner and D. L. Blackburn, Performance Trade-off for the Insulated Gate Bipolar Transistor: Buffer Layer versus Base Lifetime Reduction, *IEEE Trans. Power Electronics*, Vol. PE-2, 1987, p. 194.
13. G. Miller and J. Sack, A New Concept for a Nonpunchthrough IGBT with MOSFET like Switching Characteristics, in *Record of the Annual IEEE Power Electronics Specialists Conference (PESC1989)*, IEEE, New York, 1989, pp. 21–25.
14. T. Laska, G. Miller and J. Niedermeyer, A 2000 V Nonpunchthrough IGBT with High Ruggedness, *Solid State Electronics*, Vol. 35, No. 5, May 1992, pp. 681–685.
15. T. Laska, J. Fugger, F. Hirler, and W. Scholz, Optimizing the Vertical IGBT Structure—The NPT Concept as the Most Economic and Electrically Ideal Solution for a 1200 V IGBT, in *IEEE International Symposium on Power Semiconductor Devices and ICs (ISPSD'96)*, IEEE, New York, 1996, pp. 169–172.
16. S. K. Chung, Injection Currents Analysis of P^+/N Buffer Junction, *IEEE Trans. Electron Devices*, Vol. 45, No. 8, August 1998, pp. 1850–1854.
17. S. Pendharkar and K. Shenai, Zero Voltage Switching Behaviour of Punchthrough and Nonpunchthrough Insulated Gate Bipolar Transistors (IGBTs), *IEEE Trans. Electron Devices*, Vol. 45, No. 8, August 1998, pp. 1826–1835.
18. S. Pendharkar, M. Trivedi and K. Shenai, Electrothermal Simulations in Punchthrough and Nonpunchthrough IGBTs, *IEEE Trans. Electron Devices*, Vol. 45, No.10, October 1998, pp. 2222–2231.
19. T. P.Chow and B. J. Baliga, Comparison of N- and P-Channel IGTs, *International Electron Devices Meeting, IEDM Technical Digest*, IEEE, New York, 1984, p. 278.
20. T. P. Chow and B. J. Baliga, Comparison of 300, 600, 1200 Volt N-Channel Insulated Gate Transistors, *IEEE Electron Device Lett.*, Vol. EDL-6, 1985, pp. 161–163.
21. B. J. Baliga, M. Chang, P. Schafer and M. W. Smith, The Insulated Gate Transistor (IGT)—A New Power Switching Device, *IEEE Industrial Applications Society Meeting Digest*, IEEE, New York, 1983, pp. 794–803.
22. R. S. Ramshaw, *Power Electronics Semiconductor Switches*, Chapman & Hall, London, 1993.
23. B. J. Baliga, Switching Speed Enhancement in Insulated Gate Transistors by Electron Irradiation, *IEEE Trans. Electron Devices*, Vol. 31, No. 12, December 1984, pp. 1790–1795.
24. D. S. Kuo, J. Y. Choi, D. Giandomenico, C. Hu, S. P. Sapp, K. A. Sassaman, and R. Bregar, Modeling the Turn-off Characteristics of the Bipolar-MOS transistor, *IEEE Electron Device Lett.*, Vol. EDL-6, No. 5, May 1985, pp. 211–214.
25. A. Mogro-Campero, R. P. Love, M. F. Chang, and R. F. Dyer, Shorter Turn-off Times in Insulated Gate Transistors By Proton Implantation, *IEEE Electron Device Lett.*, Vol. EDL-6, No. 5, May 1985, pp. 224–226.
26. B. J. Baliga, Analysis of Insulated Gate Transistor Turn-off Characteristics, *IEEE Electron Device Lett.*, Vol. EDL-6, No. 2, February 1985, pp. 74–77.
27. S. Lefebvre and F. Miserey, Analysis of CIC NPT IGBTs Turn-off Operations for High Switching Current Level, *IEEE Trans. Electron Devices*, Vol. 46, No. 5, May 1999, pp. 1042–1049.
28. M. Hideshima, T. Kuramoto, and A. Nikagawa, 1000 V, 300 A Bipolar Mode MOSFET (IGBT) Module, in *Proceedings 1988 International Symposium on Power Semiconductor Devices*, 1988, pp. 80–85.

29. R. Letor, Static and Dynamic Behaviour of Paralleled IGBTs, in *Conference Record of the IEEE Industry Applications Society (IEEE-IAS) Annual Meeting*, IEEE, New York, 1990, p. 1604.
30. A. Nakagawa, Y. Yamaguchi, K. Watanabe and H. Ohashi, Safe Operating Area for 1200 V Non-latchup Bipolar-Mode MOSFETs, *IEEE Trans. Electron Devices*, Vol. ED-34, No. 2, February 1987, pp. 351–355.
31. N. Iwamuro, A. Okamoto, S. Tagami and H. Motoyama, Numerical Analysis of Short-circuit Safe Operating Area for P-Channel and N-Channel IGBTs, *IEEE Trans. Electron Devices*, Vol. 38, No. 2, February 1991, pp. 303–309.
32. H. Hagino, J. Yamashita, A. Uenishi and H. Haruguchi, An Experimental and Numerical Study on the Forward Biased SOA of IGBTs, *IEEE Trans. Electron Devices*, Vol. 43, No. 3, March 1996, pp. 490–500.
33. N. Iwamuro, A. Okamoto, S. Tagami, and H. Motoyama, Numerical Analysis of Safe Operating Area for P-Channel and N-Channel IGBTs, *IEEE Trans. Electron Devices*, Vol. 38, No. 2, February 1991, pp. 303–309.
34. I. Omura, H. Ohashi and W. Fichtner, IGBT Negative Gate Capacitance and Related Instability Effects, *IEEE Electron Device Lett.*, Vol. 18, No. 12, December 1997, pp. 622–624.
35. I. Omura, W. Fichtner and H. Ohashi, Oscillation Effects in IGBTs Related to Negative Capacitance Phenomena, *IEEE Trans. Electron Devices*, Vol. 46, No. 1, January, 1999, pp. 237–244.
36. M. Trivedi and K. Shenai, Investigation of the Short-Circuit Performance of an IGBT, *IEEE Trans. Electron Devices*, Vol. 45, No. 1, January 1998, pp. 313–320.
37. M. Trivedi and K. Shenai, IGBT Dynamics for Clamped Inductive Switching, *IEEE Trans. Electron Devices*, Vol. 45, No. 12, December 1998, pp. 2537–2545.
38. M. Trivedi and K. Shenai, Failure Mechanisms of IGBTs Under Short-circuit and Clamped Switching Stress, *IEEE Trans. Power Electronics*, IEEE, Inc., New York, Vol. 14, No. 1, January 1999, pp. 108–116.
39. B. J. Baliga, Temperature Behaviour of Insulated Gate Transistor Characteristics, *Solid State Electronics*, Vol. 28, No. 3, March 1985, pp. 289–297.
40. A. R. Hefner, D. L. Blackburn, and K. F. Galloway, The Effect of Neutrons on the Characteristics of the Insulated Gate Bipolar Transistor (IGBT), *IEEE Trans. Nuclear Science*, Vol. NS-33, No. 6, December 1986, pp. 1428–1435.
41. H. R. Chang and B. J. Baliga, 500V N-channel IGBT with Trench Gate Structure, *IEEE Trans. Electron Devices*, Vol. 36, No. 9, September 1989, pp. 1824–1829.
42. J. K. O. Sin, Lateral Trench-Gate Bipolar transistors, U.S.Patent, 5 227 653, issued July 13, 1993.
43. P. K. T. Mok, A. Nezar, and C. A. T. Salama, A Self-Aligned Trenched Cathode Lateral Insulated Gate Bipolar Transistor with High Latch-up Resistance, in *IEEE International Symposium on Power Semiconductor Devices and ICs (ISPSD'94)*, 1994, IEEE, New York, pp. 57–61.
44. D. R. Disney, H. B. Pein, and J. D. Plummer, A Trench-Gate LIGT Structure and Two LMCT Structure in SOI Substrates, in *IEEE International Symposium on Power Semiconductor Devices and ICs (ISPSD'94)*, 1994, IEEE, New York, pp. 405–410.
45. T. P. Chow, D. N. Pattanayak, B. J. Baliga, and M. S. Adler, A Reverse-Channel, High-Voltage Lateral IGBT, in *IEEE International Symposium on Power Semiconductor Devices and ICs (ISPSD'94)*, 1994, IEEE, New York, pp. 57–61.
46. F. Udrea and G. A. J. Amaratunga, A Unified Analytical Model for the Carrier Dynamics in Trench Insulated Gate Bipolar transistor, in *IEEE International Symposium on Power Semiconductor Devices and ICs (ISPSD'95)*, 1995, IEEE, New York, pp. 190–195.

47. F. Udrea and G. A. J. Amaratunga, An ON-state Analytical Model for the Trench Insulated Gate Bipolar Transistor (IGBT), *Solid State Electronics*, Vol. 41, No. 8, August 1997, pp. 1111–1118.
48. R. Sunkavalli and B. J. Baliga, Analysis of ON-State Carrier Distribution in the DI-LIGBT, *Solid State Electronics*, Vol. 41, No. 5, May 1997, pp. 733–738.
49. T. Nitta, A. Uenishi, T. Minato, S. Kusunoki, T. Takahashi, H. Nakamura, K. Nakamura, and M. Harada, A Design Concept for the Low Forward Voltage Drop 4500 V Trench IGBT, in *IEEE International Symposium on Power Semiconductor Devices and ICs (ISPSD'98)*, 1998, IEEE, New York, pp. 43–46.
50. O. Spulber, E. M. Sankara Narayanan, S. Hardikar, M. M. De Souza, M. Sweet, and S. C. Bose, A Novel Gate Geometry for the IGBT: The Trench Planar Insulated Gate Bipolar Transistor (TPIGBT), *IEEE Electron Device Lett.*, Vol. 20, No. 11, November 1999, pp. 580–582.
51. J. Cai, J. K. O. Sin, P. K. T. Mok, W.-Tung Ng, and P. P. T. Lai, A New Lateral Trench-Gate Conductivity Modulated Power transistor, *IEEE Trans. Electron Devices*, IEEE, New York, Vol. 46, No. 8, August 1999, pp. 1788–1793.
52. M. Kitagawa, S. Hasegawa, T. Inoue, A. Yahata, and H. Ohashi, 4.5 kV Injection Enhanced Gate Transistor: Experimental Verification of the Electrical Characteristics, *Japan. J. Appl. Phys.*, Vol. 36, 1997, pp. 3433–3437.
53. M. Kitagawa, A. Nakagawa, I. Omura, and J. Ohashi, Design Criterion and Operation Mechanism for 4.5 kV Injection Enhanced Gate Transistor, *Japan. J. Appl. Phys.*, Vol. 37, 1998, pp. 4294–4300.
54. H. Kon, K. Nakayama, S. Yanagisawa, J. Miwa and Y. Uetake, The 4500 V-750 A Planar Gate Press Pack IEGT, in *IEEE International Symposium on Power Semiconductor Devices and ICs (ISPSD'98)*, 1998, IEEE, New York, pp. 81–84.
55. H. Ohashi, Current and Future Development of High Power MOS Devices, *International Electron Devices Meeting, IEDM Technical Digest*, IEEE, New York, 1999, pp. 185–188.
56. Z. J. Shen, D. Briggs and S. P. Robb, Design and Characterization of High-Voltage Self-Clamped IGBTs, *IEEE Electron Device Lett.*, Vol. 20, No. 8, August 1999, pp. 424–427.
57. Z. J. Shen, D. Briggs and S. P. Robb, Voltage Dependence of Self-Clamped Inductive Switching (SCIS) Energy Capability of IGBTs, *IEEE Electron Device Lett.*, Vol. 21, No. 3, March 2000, pp. 119–122.

3

MOS COMPONENTS OF IGBT

As we have studied in the preceding chapters, the IGBT exhibits a combination of MOSFET and bipolar transistor properties. Hence, a deeper understanding of the IGBT requires a grasp of MOSFET and bipolar transistor physics. We start with the MOSFET and its associated MOS capacitor in this chapter.

3.1 GENERAL CONSIDERATIONS

3.1.1 MOS Preliminaries

Since a conducting region, termed the *channel*, exists without or with the application of a gate voltage, MOSFETs [1–7] are classified as *enhancement* or *normally-off* and *depletion* or *normally-on* devices. From the point of view of conductivity type of the channel, MOSFETs are divided into N-channel and P-channel transistors. The *N-channel enhancement-mode MOSFET* is the most commonly used structure. Figure 3.1 shows the construction features of this device. Diffusion or ion implantation produces two heavily doped N^+ regions (surface concentration $\sim 10^{19}–10^{20}$ cm^{-3}) called the *source* S and the *drain* D, separated by a distance of 1–5 μm, in a P-type Si substrate. The source is the electrode that supplies majority carriers to the channel and the drain is the electrode that collects them. Laterally, the MOSFET has the structure N^+-P-N^+. The *controlling* or *input electrode* of the MOSFET is

---

The Insulated Gate Bipolar Transistor IGBT Theory and Desig, by Vinod Kumar Khanna
ISBN: 0-471-23845-7 © 2003 by the Institute of Electrical and Electronics Engineers

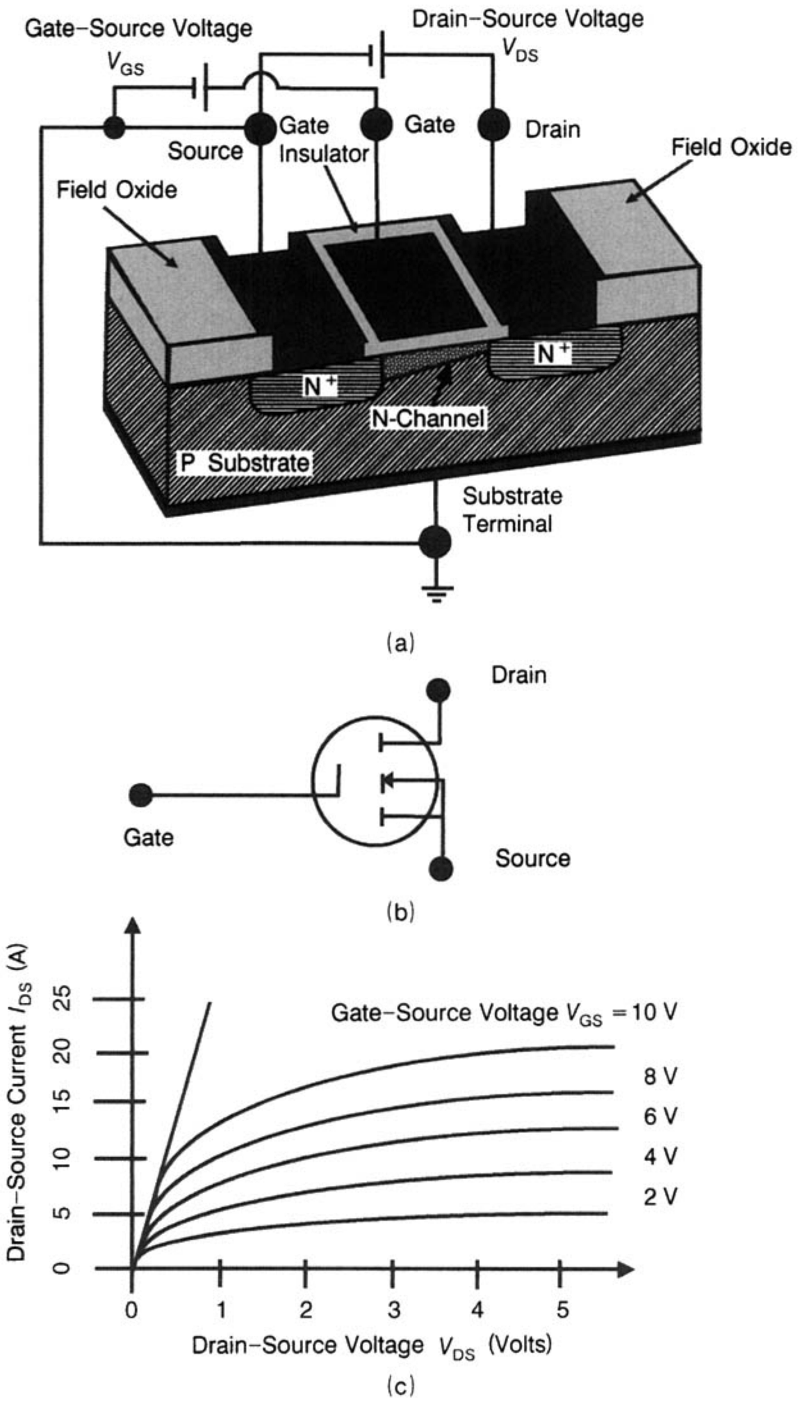

Gate–Source Voltage
V_{GS}
Drain–Source Voltage
V_{DS}
Source
Gate
Insulator
Gate
Drain
Field Oxide
Field Oxide
N+
N+
N-Channel
P Substrate
Substrate
Terminal
(a)
Drain
Gate
Source
(b)
Drain–Source Current I_{DS} (A)
0
5
10
15
20
25
Gate–Source Voltage V_{GS} = 10 V
8 V
6 V
4 V
2 V
0
1
2
3
4
5
Drain–Source Voltage V_{DS} (Volts)
(c)

Figure 3.1 N-channel enhancement-mode MOSFET: (a) Schematic cross section, (b) circuit symbol, and (c) typical output characteristics.

known as the *gate* G covering the space between source and drain and isolated from the semiconductor by an intervening dielectric layer (SiO_2, SiO_2–Si_3N_4, or SiO_2–Al_2O_3) providing an input resistance $\sim 10^{12}$–10^{15} Ω. Apart from the source, drain, and gate, the MOSFET has a fourth terminal B, termed the *body* or *substrate* terminal, so that the MOSFET is a four-terminal device.

When the applied gate-source voltage V_{GS} is zero, source and substrate terminals are grounded together ($V_{SB} = 0$) and the drain-source voltage V_{DS} is positive; only a small current $\sim 10^{-10}$–10^{-15} A, corresponding to a reverse-biased PN junction, flows from source to drain. Now suppose the

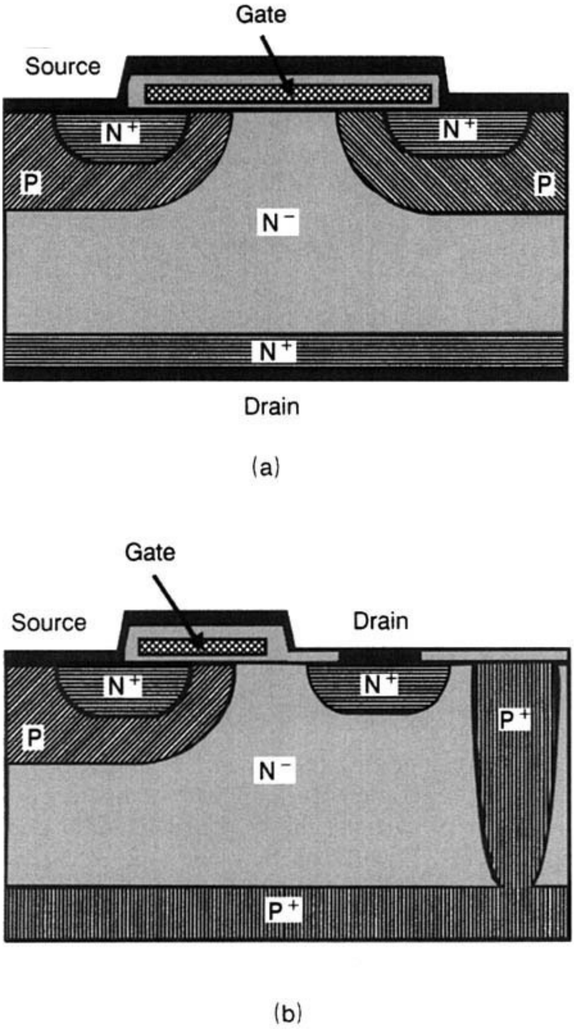


Figure 3.2 Power MOSFET structures: (a) Vertical and (b) lateral.

gate-source voltage V_{GS} is gradually increased from zero value. Appreciable current starts flowing between source and drain at a certain V_{GS} value exceeding the threshold voltage rating V_{Th} of the device, that is $V_{GS} > V_{Th}$. How this increase of V_{GS} above the threshold voltage influences the current flow in the Si below is understood by a detailed analysis of the MOS structure in Section 3.2.

3.1.2 Power MOSFET Structures

The advent of power MOSFETs [3, 4], especially with a vertical geometry (Fig. 3.2a), ranks as a landmark in power electronics. Power MOSFETs have a vertical structure so that current flows across the pellet between the power terminals on opposite sides, yielding low ON-state voltage drop and high current capability. However, this voltage drop is higher than that of equivalent junction transistor. Power MOSFETs, with typical current and voltage ratings of tens of amperes and hundreds of volts, are available, which are much lower than the corresponding ratings of junction transistors. The lateral power MOSFET having a planar structure (Fig. 3.2b) is employed in power ICs. In this structure, P-N junction isolation is used (see Section 9.3).

There are three popular power vertical-channel discrete MOSFET configurations: the V-groove or VMOSFET (Fig. 3.3a), the double-diffused or DMOSFET (Fig. 3.3b), and U-groove or UMOSFET (Fig. 3.3c), also called trench–gate MOSFET. The salient features of these three structures are presented in Table 3.1.

3.1.3 MOSFET-Bipolar Transistor Comparison

A built-in diode called the "body diode" exists in the internal junction structure of a power MOSFET, making it a parallel combination of two static switches, namely, a controlled MOSFET switch for forward currents and an uncontrolled diode switch for reverse currents (Fig. 3.4). This combination of switches is extremely useful for static converters where the diode provides a free-wheeling current path. Since the current and switching speed of this diode is sufficient for many applications, an externally connected freewheeling diode is hardly required. But junction power transistors usually need proper fast recovery diodes.

Recapitulating and elaborating the MOSFET and bipolar transistor comparison in Table 1.2, operation of a bipolar transistor is based on the transport of two types of carriers, minority and majority carriers, electrons and holes. But a MOSFET works on only one type of carrier, the majority carriers, either electrons or holes, hence it is also called a *unipolar device*. Also, the bipolar transistor is a current-controlled device whereas the MOSFET is a voltage-controlled device, based on the field-effect principle utilizing the surface conductivity modulation of a semiconductor, by a transverse electric field, in the longitudinal direction. Another major difference between the two types of devices is that the applied base drive during forward

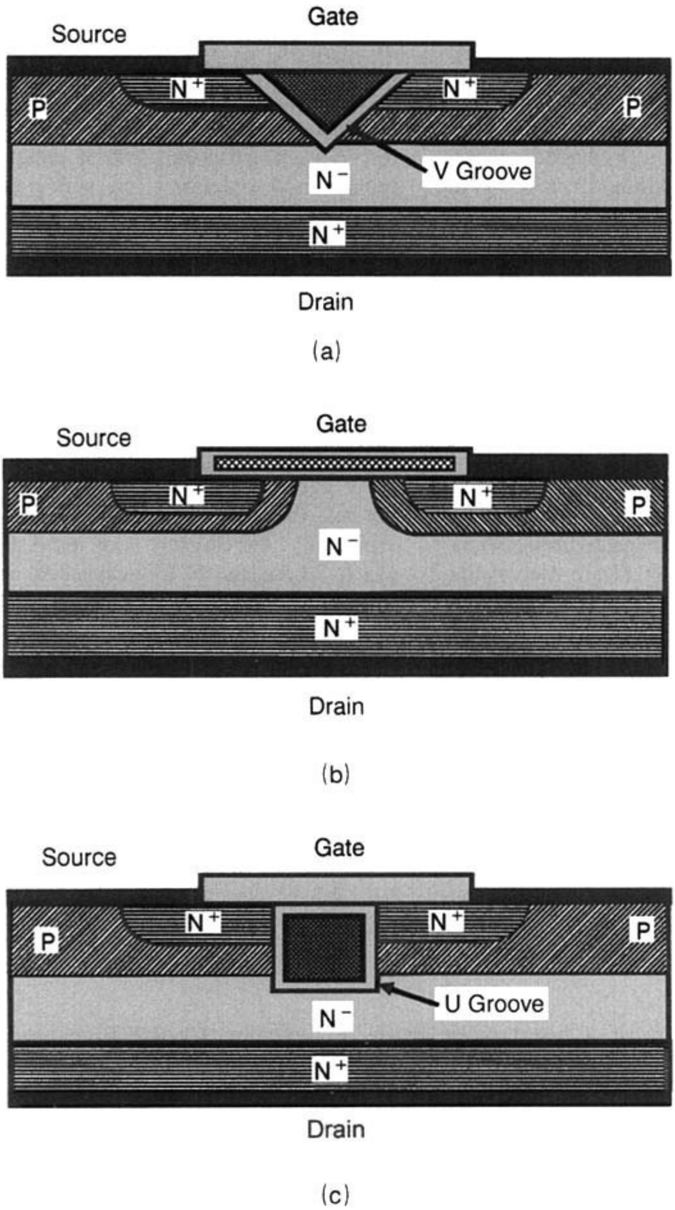

Source
Gate
N^+
N^+
P
P
N^-
V Groove
N^+
Drain
(a)
Source
Gate
P
N^+
N^+
P
N^-
N^+
Drain
(b)
Source
Gate
P
N^+
N^+
P
U Groove
N^-
N^+
Drain
(c)

Figure 3.3 Common MOSFET structures: (a) VMOSFET, (b) DMOSFET, and (c) UMOSFET.

Table 3. 1 Vertical Discrete MOSFET Structures

Serial No.	VMOSFET	DMOSFET	MOSFET
1	Derives its name from the V-shaped groove within which the gate is situated.	Derives its name from the double-diffusion process used to define the channel length. Difference of lateral diffusion between P-base and N^+ source gives the channel length.	Derives its name from the U-shaped groove in which the gate is located. This groove is made by reactive ion etching (RIE).
2	Fabricated by performing an unpatterned P-base diffusion and then N^+ diffusion followed by V-grooving by preferential etching. Gate deposition is done and photolithography is carried out to overlap the N^+ source and extend it into the groove beyond P-base bottom.	Fabricated by planar techniques using a polySi refractory gate as a mask. Its edge defines the common window through which P-base and N^+ source regions are sequentially diffused and driven-in.	Fabrication process sequence is similar to VMOSFET. The technology is based on trench etching methods used for storage capacitor in memories.
3	Stability problems are faced during manufacturing because a high local electric field is created at the V-groove tip.	Superseded VMOSFET devices, commercially.	Provides highest channel density of the three varieties giving appreciably lower ON resistance.

conduction of a bipolar transistor is typically 0.1–0.2 times the collector current and even higher reverse currents are required for turn-off, making the base drive circuits complex and costly. In a MOSFET, gate current pulses of relatively smaller magnitude and duration are necessary to charge and discharge the gate capacitance.

Furthermore, the bipolar transistor is prone to second breakdown, as when a high collector–emitter voltage is applied, in the presence of a high collector current, commonly encountered during inductive load turn-off. This happens due to the nonuniformity of current distribution resulting from material inhomogeneities or design limitations, whereby current flow is constricted in certain regions such as the centers of emitter regions. The device fails via the negative temperature coefficient of forward resistance, whereby current increases with temperature result in hot spot formation, leading to microplasmas and mesoplasmas. Bipolar transistors therefore

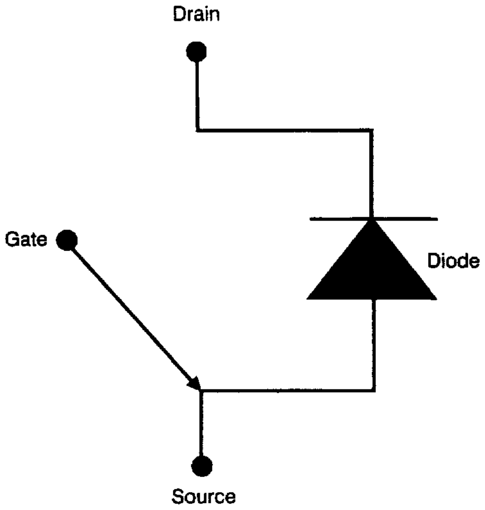


Figure 3.4 Functional representation of MOSFET as a combination of two switches.

cannot be operated in parallel unless some means of emitter ballasting (e.g., resistors) are used to prevent the shoot through in current. This fundamental drawback severely restricts the safe operating area of a bipolar transistor. On the other hand, MOSFET operation is not hampered by a second breakdown limitation, if the turning on of the parasitic bipolar transistor is averted. Owing to its positive temperature coefficient of ON resistance, it displays a larger safe operating area than the bipolar transistor. Several MOSFET devices are easily connected in parallel, with proper use of gate drive resistors, to increase the current-carrying capability. Finally, due to absence of minority-carrier storage phenomenon in MOSFET, this device exhibits a very fast response. It is the fastest power semiconductor switch, useful in the high-frequency, relatively low-power dominion.

3.2 MOS STRUCTURE ANALYSIS AND THRESHOLD VOLTAGE

The ideal MOS structure is defined as one fulfilling the following criteria: (i) Resistivity of the insulating SiO_2 layer is infinitely large. (ii) Charge can exist only on the metal electrode and in the Si. (iii) Work functions of the metal and semiconductor are the same; interface charge, fixed oxide charge, and so on, are absent.

Figure 3.5a shows the energy-band diagram of an ideal MOS structure with P-type semiconductor when the applied gate voltage is zero. Here, ϕ_M = work function of the metal = minimum amount of energy, measured in joules or electron volts, that must be imparted to the fastest moving electron at 0 K in order that the electron escapes from the metallic surface, ϕ_B =

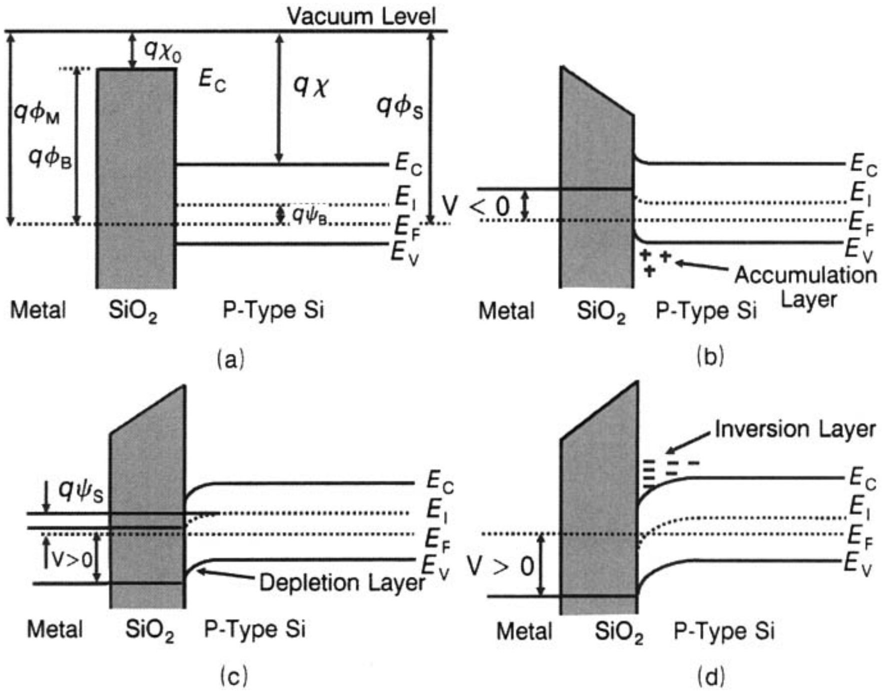

Vacuum Level
qχ0
qφM
qφB
EC
qχ
qφS
EC
EI
qψB
EF
EV
Metal
SiO2
P-Type Si
(a)
V < 0
EC
EI
EF
EV
Accumulation Layer
Metal
SiO2
P-Type Si
(b)
qψS
V>0
EC
EI
EF
EV
Depletion Layer
Metal
SiO2
P-Type Si
(c)
Inversion Layer
V > 0
EC
EI
EF
EV
Metal
SiO2
P-Type Si
(d)

Figure 3.5 Energy band diagrams of MOS structure under different gate biasing conditions: (a) Without any gate bias, (b) with a negative gate bias, (c) with a small positive gate bias, and (d) with a large positive gate bias.

barrier height between metal, and SiO_2 = energy that must be surmounted by the electrons to move from the metal into the SiO_2, χ_0 = electron affinity of SiO_2 = energy released in electron volts (eV) when a molecule of SiO_2 gains an electron to form a negative ion, χ = electron affinity of Si (defined for the Si atom in the same manner as done for SiO_2 above), ψ_B = potential difference between intrinsic energy level E_I and the Fermi level E_F in Si. As can be seen from the band diagram, there is no band bending. This is known as the *flatband condition.* Under this condition, the following equation holds [5–7]

$$q\phi_M = q\chi + \frac{E_g}{2} + q\psi_B = q(\phi_B + \chi_0) \tag{3.1}$$

This equation is readily understood by recalling the definitions of the work function, electron affinity, and barrier height. The work function $q\phi_M$ = energy difference between the Fermi level E_F and the vacuum level = energy difference between the bottom of the conduction band, E_C in Si, and the vacuum level (electron affinity of Si = $q\chi$) + energy difference between the intrinsic energy level, E_I in Si, and conduction band bottom E_C (~ half of the energy band gap E_g) + energy difference separating the intrinsic energy level E_I in Si from the Fermi energy level E_F (given by $q\psi_B$) = energy difference between the Fermi level E_F and the bottom of the conduction band in SiO_2 (barrier height of $SiO_2 = q\phi_B$) + energy difference between the conduction band bottom and the reference vacuum level in SiO_2 (electron affinity of $SiO_2 = q\chi_0$).

In Fig. 3.5b, a negative gate voltage is applied to the structure. Attraction of holes toward the Si surface bends the valence band closer to the Fermi level near the surface to allow for the states occupied by these holes. This condition is called *accumulation.* Figure 3.5c illustrates the case of application of a small positive gate voltage. Now holes are repelled away from the surface toward the interior. As the region adjoining the Si surface is depleted of free charge carriers, the valence band curves away from the Fermi level, indicating that the probability of occupancy of states near the surface by holes, is decreased. Consequently, a depletion layer is created in the vicinity of the surface. This condition is known as *depletion.* Figure 3.5d depicts the situation when the positive gate bias is so large that the intrinsic energy level crosses the Fermi level. Then the concentration of minority-carrier electrons becomes greater than the majority-carrier hole concentration. This condition is called *inversion* and leads to the formation of an *inversion layer* having N-type conductivity and is about 10 nm thick, playing the central role in current transport in the MOSFET. Inversion layer charge determination requires knowledge of potential distribution $\psi(x)$ in the semiconductor. If n_{p0} and p_{p0} denote thermal equilibrium electron and hole concentrations in the bulk semiconductor, their concentrations at the surface are given by (Appendix 3.1)

$$n_p = n_{p0} \exp\left(\frac{q\psi}{kT}\right) \tag{3.2a}$$

and

$$p_p = p_{p0} \exp\left(-\frac{q\psi}{kT}\right) \tag{3.2b}$$

where the potential distribution $\psi(x)$ obeys the Poisson's equation

$$\frac{d^2\psi}{dx^2} = -\frac{\rho(x)}{\varepsilon_0 \varepsilon_s} \tag{3.3}$$

In this equation, ε_0 and ε_s symbolize the permittivity of free space and relative permittivity of Si, respectively, and ρ is the charge density written as

$$\rho(x) = q(N_{D+} - N_{A-} + p_p - n_p) \tag{3.4}$$

where N_{D+}, N_{A-} are the donor and acceptor ion concentrations. The boundary condition of existence of charge neutrality in the bulk semiconductor far away from the surface can be expressed as

$$N_{D+} - N_{A-} = n_{p0} - p_{p0} \tag{3.5}$$

Using equations for n_p and p_p along with the charge-neutrality condition allows us to recast Poisson's equation in the form

$$\frac{d^2\psi}{dx^2} = -\frac{q}{\varepsilon_0 \varepsilon_\mathrm{s}}\left[p_{\mathrm{p}0}\left\{\exp\left(-\frac{q\psi}{kT}\right) - 1\right\} - n_{\mathrm{p}0}\left\{\exp\left(\frac{q\psi}{kT}\right) - 1\right\}\right] \tag{3.6}$$

Integration of this equation from the bulk toward the surface gives the electric field distribution $\xi(x)$ as (Appendix 3.2)

$$\xi(x) = -\frac{d\psi}{dx}$$

$$= \frac{\sqrt{2}kT}{qL_\mathrm{D}}\left[\exp\left(-\frac{q\psi}{kT}\right) + \frac{q\psi}{kT} - 1 + r\left\{\exp\left(\frac{q\psi}{kT}\right) - \frac{q\psi}{kT} - 1\right\}\right]^{0.5} \tag{3.7}$$

where $r = n_{\mathrm{p}0}/p_{\mathrm{p}0}$ and $L_\mathrm{D} = \sqrt{\left\{kT\varepsilon_0\varepsilon_s/\left(q^2 p_{\mathrm{p}0}\right)\right\}}$ is the intrinsic Debye length for holes. *Debye length* is a characteristic length, also called the *shielding* or *screening length*, measured by the distance over which the electric field emanating from a perturbing charge falls off by a factor of $1/e = 0.37$. It gives an estimate of the distance in which a charge imbalance is smeared out. As an example, if a negatively charged sphere is introduced into a P-type semiconductor, the free holes in the semiconductor will crowd around the sphere. Several Debye lengths away from the sphere, the negatively charged sphere together with the positively charged hole cloud will look like a neutral object.

Now applying Gauss' law, the space charge Q_s required per unit area to produce this electric field is

$$Q_\mathrm{s} = -\varepsilon_0 \varepsilon_\mathrm{s} \xi_\mathrm{s} \tag{3.8}$$

where ξ_s is the electric field at the surface. If the potential at the surface is ψ_s, the surface charge $(Q_\mathrm{s})_\mathrm{surface}$ is obtained by replacing ψ by ψ_s in the equation for ξ.

When the surface potential ψ_s increases from zero toward the positive side, three pertinent cases arise:

Case I: Surface Potential ψ_s < Bulk Potential ψ_B: The intrinsic energy level E_I does not cross the Fermi level E_F. The surface polarity is P-type containing a depletion layer. The second term in Eq. (3.7) $q\psi/(kT)$ dominates and the surface charge becomes

$$Q_\mathrm{s} = \varepsilon_0 \varepsilon_\mathrm{s} \xi_\mathrm{s} = \varepsilon_0 \varepsilon_\mathrm{s} \frac{\sqrt{2}kT}{qL_\mathrm{D}} \times \left(\frac{q\psi_\mathrm{s}}{kT}\right)^{0.5} = \varepsilon_0 \varepsilon_\mathrm{s} \frac{\sqrt{2}kT}{q\sqrt{\dfrac{kT\varepsilon_0\varepsilon_\mathrm{s}}{q^2 p_{\mathrm{p}0}}}} \times \sqrt{\frac{q\psi_\mathrm{s}}{kT}}$$

or

$$Q_s = \sqrt{2\varepsilon_0 \varepsilon_s q p_{p0} \psi_s} \tag{3.9}$$

where the equation for L_D has been used. Thus the surface charge is proportional to the square root of the surface potential.

Case II: $\psi_s > \psi_B$ but $\psi_s < 2\psi_B$ (Weak Inversion): A mobile carrier concentration builds up at the surface. The regime $\psi_s < 2\psi_B$ is known as *weak inversion.*

Case III: $\psi_s > 2\psi_B$ (Strong Inversion): The fourth term in Eq. (3.7) becomes dominant, giving

$$Q_s = \varepsilon_0 \varepsilon_s \xi_s = \varepsilon_0 \varepsilon_s \frac{\sqrt{2}kT}{qL_D} \times \left\{ r \exp\left(\frac{q\psi_s}{kT}\right)\right\}^{0.5}$$

$$= \varepsilon_0 \varepsilon_s \frac{\sqrt{2}kT}{q\sqrt{\dfrac{kT\varepsilon_0\varepsilon_s}{q^2 p_{p0}}}} \times \left\{ \frac{n_{p0}}{p_{p0}} \exp\left(\frac{q\psi_s}{kT}\right)\right\}^{0.5}$$

or

$$Q_s \cong \sqrt{2\varepsilon_0 \varepsilon_s n_{p0}} \exp\left(\frac{q\psi_s}{2kT}\right) \tag{3.10}$$

that is, the surface charge varies exponentially with surface potential. This charge controls the channel conductivity of the MOSFET.

The parameter of supreme importance controlling the operation of MOS devices is the *threshold voltage* V_{Th}, which is defined as the minimum bias applied to the gate electrode at which an N-type conducting channel appears at the Si surface, producing a large current flow between source and drain, as evidenced by the onset of *strong inversion.* An optimum value of V_{Th} is called for, neither too small nor excessively large. Too small a threshold voltage may result in a normally-on device, due to the positive gate oxide charge. Even if the MOSFET is not always on, a low-threshold-voltage device is plagued by inadvertent false turn-on due to noise spikes at the gate or temperature effects. During high-speed switching, the gate voltage may be forced to remain continuously fixed at values above threshold voltage rendering the device always conducting. On the opposite extreme, higher threshold voltages are also undesirable because the supply voltage must be increased and the input power is raised. Furthermore, a MOSFET having a high threshold voltage is incompatible with logic circuits due to interfacing problems. This may unnecessarily complicate the design of gate driving circuits and warrant the use of buffer elements. Generally, the threshold voltage of a power

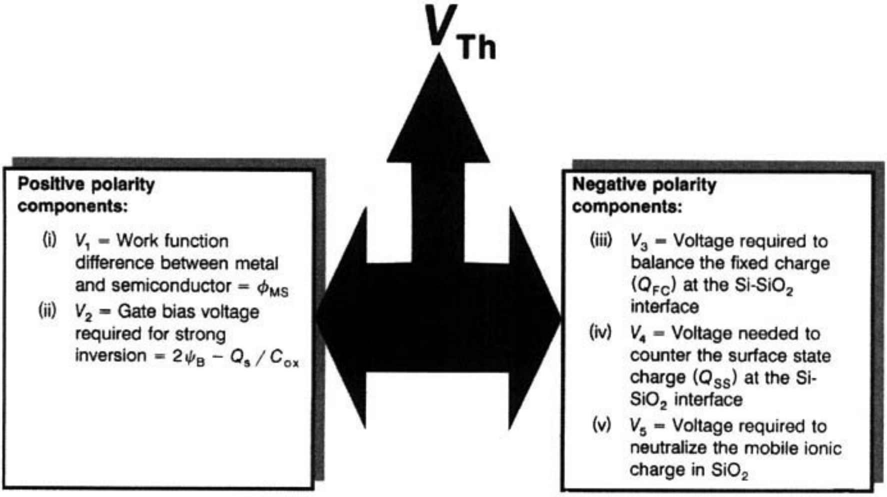

V_{Th}
Positive polarity components:
(i) V_1 = Work function difference between metal and semiconductor = ϕ_{MS}
(ii) V_2 = Gate bias voltage required for strong inversion = $2\psi_B - Q_s / C_{ox}$
Negative polarity components:
(iii) V_3 = Voltage required to balance the fixed charge (Q_{FC}) at the Si-SiO$_2$ interface
(iv) V_4 = Voltage needed to counter the surface state charge (Q_{SS}) at the Si-SiO$_2$ interface
(v) V_5 = Voltage required to neutralize the mobile ionic charge in SiO$_2$

Figure 3.6 Components of the threshold voltage of N-channel MOSFET.

MOSFET is designed to lie between 2 and 4 V. Some applications require better switching performance, necessitating interface driver circuits.

V_{Th} has the following components (Fig. 3.6):

(i) $V_1 = \phi_{MS}$, the work function difference between metal (Al) and Si = the contact potential difference. If ϕ_B is the barrier height between metal and SiO$_2$, we can write

$$\phi_{MS} = \text{Work function of metal } (\phi_M) - \text{Work function of semiconductor } (\phi_S)$$

or

$$q\phi_{MS} = q\phi_B + q\chi_0 - q\left(\chi + \frac{E_g}{2q} + \psi_B\right) \tag{3.11}$$

This equation readily follows from the inspection of Fig. 3.5a, recalling the definitions of work function (ϕ), barrier height (ϕ_B), and electron affinity (χ), and the discussion in connection with Eq. (3.1).

(ii) $V_2 = 2\psi_B - Q_s/C_{ox}$ is the voltage required to establish the inversion layer—that is, to create the depletion layer and shift the minority carriers in the substrate toward the surface.

(iii) $V_3 = Q_{FC}/C_{ox}$, where Q_{FC} is the fixed charge at the Si–SiO$_2$ interface (Fig. 3.7) and C_{ox} is the specific capacitance (capacitance per unit area)

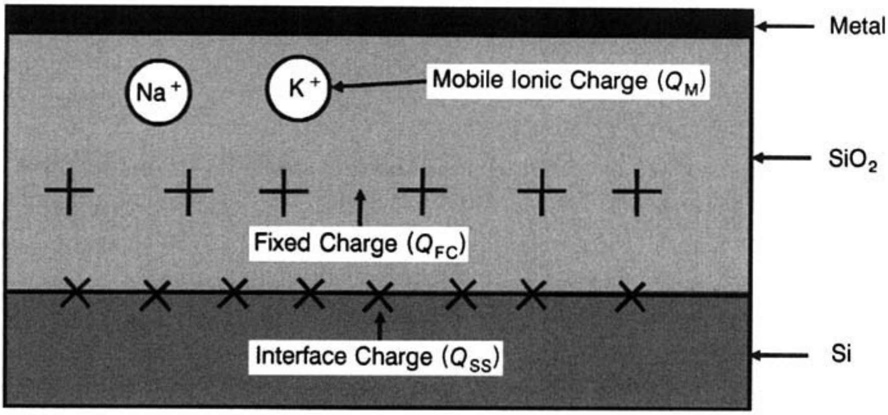


Figure 3.7 Different types of charge in the MOS system.

of the oxide given by

$$C_{ox} = \frac{\varepsilon_0 \varepsilon_{ox}}{d} \tag{3.12}$$

where ε_{ox} is the dielectric constant of SiO$_2$ and d is the thickness of gate oxide. The charge Q_{FC} lies within 20 Å of the Si–SiO$_2$ interface. Its value depends on the oxide growth conditions and its subsequent heat treatment, and this is attributed to oxygen deficiency or excess Si present. It is neither chargeable nor dischargeable at normal device potentials.

(iv) $V_4 = Q_{SS}/C_{ox}$, where Q_{SS} is the interface-state charge (Fig. 3.7) located in the vicinity of the surface with energy levels in the band gap. The interface states are produced by the dangling or unsaturated bonds at the surface and have a density around 10^{11}–10^{12} cm^{-2}-eV. This density is dependent on the orientation of the Si crystal with $\langle 100 \rangle$ orientation providing an order of magnitude lower density than $\langle 111 \rangle$ orientation. Hydrogen annealing at 450°C effectively neutralizes Q_{SS}. These states are charged and discharged by the applied surface potential. They cause deviations in the capacitance–voltage or channel conductance–voltage characteristics of MOS structures from the ideal theoretical curve shapes.

(v) $V_5 = Q_M/C_{ox}$, where Q_M is the mobile ion charge due to Na$^+$ ions, or other mobile ionic species, which cause drift in the threshold voltage under bias stress at high temperatures over extended periods. Their concentration is determined by the processing cleanliness.

Adding together the above components, we have

$$V_{\mathrm{Th}} = \phi_{\mathrm{MS}} + 2\psi_{\mathrm{B}} - \frac{Q_{\mathrm{s}} + Q_{\mathrm{FC}} + Q_{\mathrm{SS}} + Q_{\mathrm{M}}}{C_{\mathrm{ox}}} \tag{3.13}$$

Substituting for Q_s at the point of transition into strong inversion and for ψ_B from Appendix 3.3, and for C_{ox} from Eq. (3.12), the V_{Th} equation may be written as

$$V_{\mathrm{Th}} = \phi_{\mathrm{MS}} + \frac{2kT}{q}\ln\frac{N_{\mathrm{A}}}{n_{\mathrm{i}}} - \frac{\sqrt{4\varepsilon_0\varepsilon_{\mathrm{s}}kTN_{\mathrm{A}}\ln(N_{\mathrm{A}}/n_{\mathrm{i}})}}{\varepsilon_0\varepsilon_{\mathrm{ox}}t_{\mathrm{ox}}} - \frac{Q_{\mathrm{FC}} + Q_{\mathrm{SS}} + Q_{\mathrm{M}}}{C_{\mathrm{ox}}} \tag{3.14}$$

Due to different barrier heights of metal on SiO_2, the threshold voltage of a MOSFET can be varied, within a limited range, by proper choice of gate metal. More precise control of threshold voltage is afforded by substituting the metallic gate electrode with a heavily doped polycrystalline silicon gate electrode. The work function difference of the polySi–SiO_2–monoSi system is a function of the doping concentration of the polySi layer. Due to the difference in the doping concentrations of N-type polySi gate and the underlying P-type monocrystalline silicon, a contact potential develops between them in exactly the same manner as across an ordinary P-N junction. At the normal doping levels encountered in MOS devices, the magnitude of the contact potential is the same as obtained at a P-N junction formed from the same materials. Remembering that the polarity of electrostatic potential is opposite to that of electron energy, the N^+ polySi gate is positive with respect to the P-type silicon below. If N_A is the doping concentration of the monocrystalline P-type substrate and N_{DpolySi} that of the polySi layer, the work function difference polySi/monoSi is written as

$$q\phi_{\mathrm{polySi\text{-}monoSi}} = -\frac{kT}{q}\ln\left(\frac{N_{\mathrm{DpolySi}}\,N_{\mathrm{A}}}{n_{\mathrm{i}}^2}\right) \tag{3.15}$$

For derivation of the equation of built-in potential of a P-N junction, the reader may refer to Section 4.1.1, Eq. (4.6).

Polysilicon has the advantage of being a refractory material compatible with silicon device processing unlike the metal gate electrode, which cannot be subjected to high temperature treatments. The disadvantage is its lower conductivity, about 10 times less than the metal electrode, which makes the RC gate time charging time constant too large at high frequencies. For such applications, Al is preferred.

Example 3.1 Calculate the flatband voltage (V_{FB}) of the N-channel MOS structure in an IGBT with Al gate having P-base concentration = 2×10^{17} cm^{-3}, gate oxide thickness = 500 Å, and interface charge $Q_0 = 1 \times 10^{-8}$ C/cm^2. What is the flat-band voltage if a heavily doped polySi gate ($N_D = 1 \times 10^{20}$ cm^{-3}) replaces the Al gate? Contact potential of Al with intrinsic Si is 0.6 V.

Flat-band voltage V_{FB} is the external voltage applied between the gate and P-base terminals to maintain the semiconductor everywhere neutral. To derive the equation for the flatband voltage, let us consider the energy band diagram of the MOS structure shown in Fig. E3.1.1. When the metal, oxide and semiconductor are joined

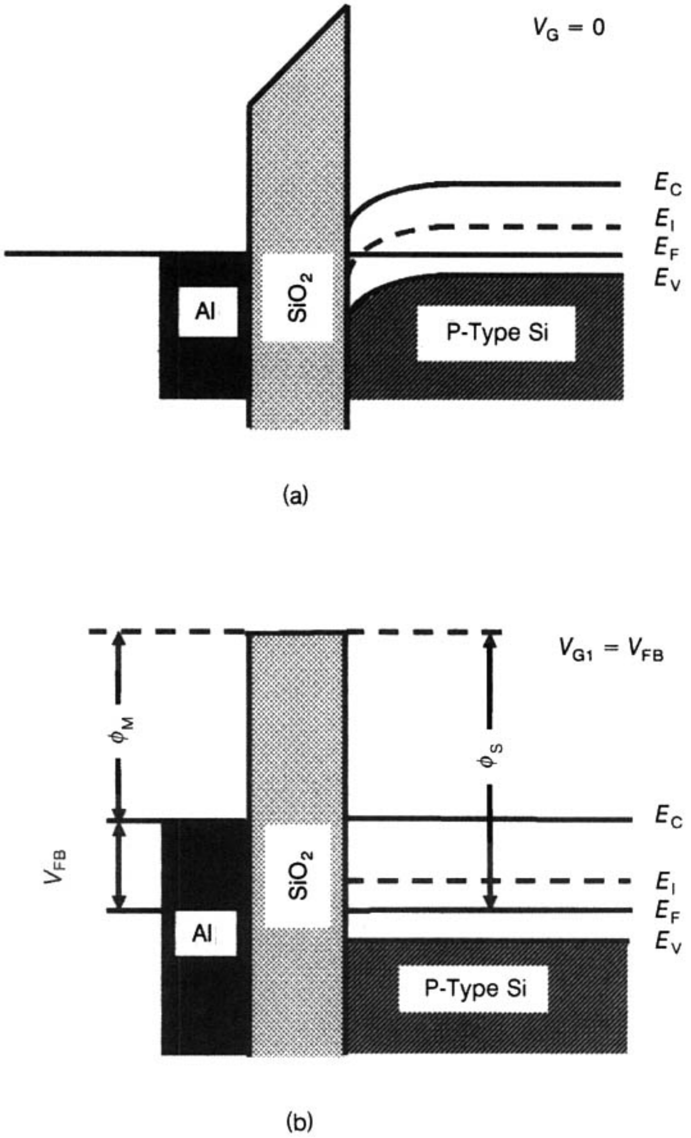

$V_G = 0$
E_C
E_I
E_F
E_V
Al
SiO$_2$
P-Type Si
(a)
$V_{G1} = V_{FB}$
ϕ_M
ϕ_S
V_{FB}
E_C
E_I
E_F
E_V
Al
SiO$_2$
P-Type Si
(b)

Figure E3.1.1 Energy band diagrams of the Al–SiO$_2$–Si structure: (a) For $V_G = 0$ V and (b) under flat-band condition.

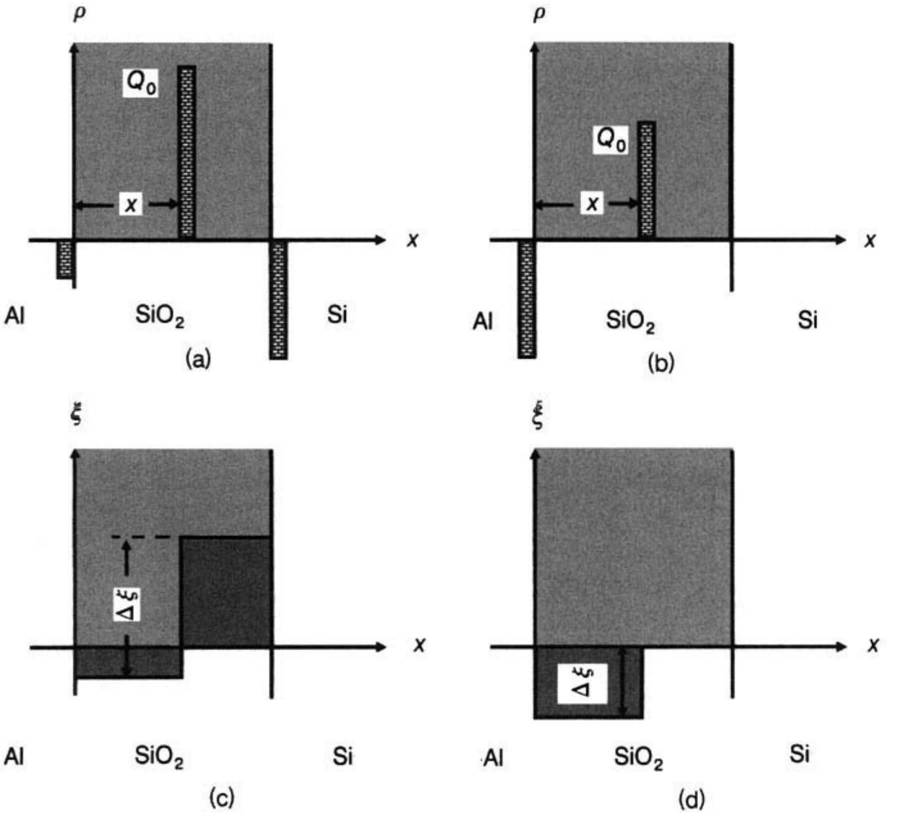


Figure E3.1.2 Effect of a sheet charge in the oxide: (a) Charge distribution for $V_G = 0$, (b) resulting electric field, (c) charge distribution for flat-band condition, and (d) corresponding electric field.

together to form the MOS structure, a band bending takes place to satisfy the requirement of a constant Fermi level under thermal equilibrium. In place of the metal and semiconductor work functions, the appropriate energies to be considered are their modified work functions, measured from the respective Fermi level in the metal and the semiconductor to the oxide conduction band edge. The gate voltage V_{GI} required to counterbalance the modified work function difference (so that no electric field exists inside it), and maintain a flat-band condition in the semiconductor, is clearly the difference of the modified work functions ϕ_{MS}:

$$V_{GI} = \phi_{MS} \tag{E3.1.1}$$

This gate voltage is further modified by the charges residing within the oxide. With reference to Fig. E3.1.2a, let us consider a thin layer of charge $+Q_0$ per unit area located at a distance x from the metal. Under zero gate voltage condition, this sheet charge will induce an image charge, partly in the metal and partly in the semiconductor, and the sum of the induced charges $= -Q_0$. The electric field distribution obtained by integrating the charge distribution, is plotted in Fig. E3.1.2b. The induced charge in the semiconductor produces a nonzero electric field at its surface, disturbing the flat energy-band diagram. Band flattening is accomplished by applying a negative voltage to shift the electric field distribution downward so that the electric field reaching the semiconductor surface, and hence charge becomes zero. The electric field increment is

$$\Delta \xi = \int_0^x \frac{\rho\, dx}{\varepsilon_0 \varepsilon_{\text{ox}}} = \frac{\rho x}{\varepsilon_0 \varepsilon_{\text{ox}}} = \frac{Q_0}{\varepsilon_0 \varepsilon_{\text{ox}}} \tag{E3.1.2}$$

where ρ is the charge density per unit volume, ε_0 is the permittivity of free space and ε_{ox} is the dielectric constant of silicon dioxide. The charge distribution is sketched in Fig. E3.1.2c, and the electric field distribution in Fig. E3.1.2d. Then the gate voltage required to achieve flat-band condition is

$$V_{\text{G2}} = -\int_0^x \Delta \xi\, dx = -\frac{xQ_0}{\varepsilon_0 \varepsilon_{\text{ox}}} = -\frac{x}{x_{\text{ox}}} \frac{Q_0}{C_{\text{ox}}} \tag{E3.1.3}$$

where x_{ox} is the oxide thickness and C_{ox} is oxide capacitance per unit area. This equation shows that V_{G2} depends not only on the density of the charge (Q_0) but also on its location within the insulator. V_{G2} is zero if the charge is situated at $x = 0$. On the other hand, the charge exerts its maximum influence when it is located at $x = x_0$ —that is, at the oxide–silicon interface. Since the oxide charge is generally found at this interface, $x/x_0 = 1$ and Eq. (E3.1.3) reduces to

$$V_{\text{G2}} = -\frac{Q_0}{C_0} \tag{E3.1.4}$$

Adding together V_{G1} and V_{G2} [Eqs. (E3.1.1) and (E3.1.4)], the flat-band voltage is given by

$$V_{\text{FB}} = \phi_{\text{MS}} - \frac{Q_0}{C_{\text{ox}}} \tag{E3.1.5}$$

where ϕ_{MS} is the modified metal–semiconductor work function, Q_0 is the effective interface charge per unit area and C_{ox} is the oxide capacitance per unit area. Figure E3.1.3a shows the unit cell of the Al-gate IGBT. Here $\phi_{\text{MS}} = \phi_{\text{bulk material}} - \phi_{\text{gate material}} = -\phi_{\text{F}} - \phi_{\text{i}}$ where ϕ_{F} is the Fermi potential of extrinsic Si (contact potential between intrinsic and extrinsic Si) and ϕ_{i} is the contact potential of material i to intrinsic Si. $\phi_{\text{F}} = +(kT/q)\ln(N_{\text{A}}/n_{\text{i}}) = 0.0259\ln(2 \times 10^{17}/1.45 \times 10^{10}) = 0.4258$ V. Since $\phi_{\text{i}} = 0.6$ V, $\phi_{\text{MS}} = -0.4258 - 0.6 = -1.0258$ V. Now $C_{\text{ox}} = \varepsilon_0 \varepsilon_{\text{ox}}/d_{\text{ox}} = 8.854 \times 10^{-14} \times 3.9/500 \times 10^{-8} = 6.906 \times 10^{-8}$ F/cm^2. So, $Q_0/C_{\text{ox}} = 1 \times 10^{-8}/6.906 \times 10^{-8} = 0.1448$ and $(V_{\text{FB}})_1 = -0.966 - 0.1448 = -1.1108$ V.

For degenerately doped polySi gate IGBT (Fig. E3.1.3b) $q\phi_{\text{pSi-monoSi}} = (kT/q)\ln(N_{\text{DpSi}} N_{\text{A}}/n_{\text{i}}^2) - E_{\text{g}} = 0.0259\ln\{1 \times 10^{20} \times 2 \times 10^{17}/(1.45 \times 10^{10})^2\} - 1.11 = -0.9747$. Hence $(V_{\text{FB}})_2 = -0.9747 - 0.1448 = -0.24227$. Thus $(V_{\text{FB}})_2 - (V_{\text{FB}})_1 = -0.24227 + 1.1108 = 0.86853$ V.

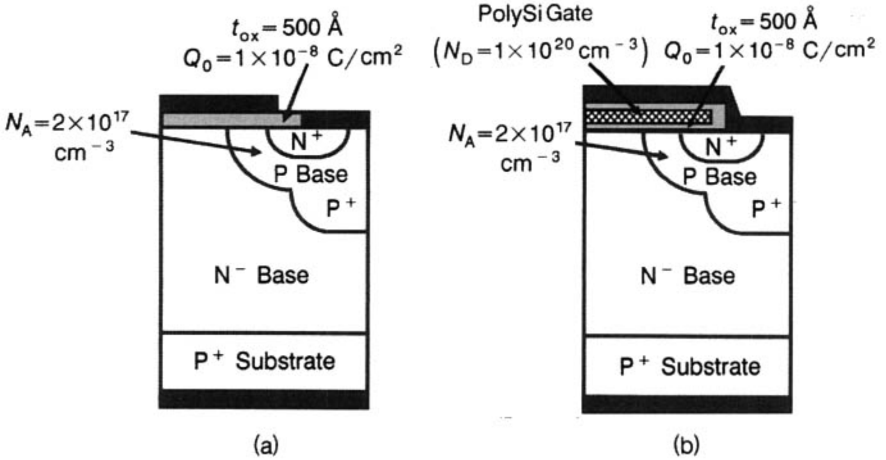

t_{ox} = 500 Å
Q_0 = 1 × 10^{-8} C/cm²
N_A = 2 × 10^{17} cm⁻³
N⁺
P Base
P⁺
N⁻ Base
P⁺ Substrate
(a)
PolySi Gate
(N_D = 1 × 10^{20} cm⁻³)
t_{ox} = 500 Å
Q_0 = 1 × 10^{-8} C/cm²
N_A = 2 × 10^{17} cm⁻³
N⁺
P Base
P⁺
N⁻ Base
P⁺ Substrate
(b)

Figure E3.1.3 IGBT half-cell structures and physical parameters for Example 3.1: (a) Al gate and (b) polySi gate.

Example 3.2 (a) Find the threshold voltage of an N-channel IGBT having an N^+ silicon gate if the gate oxide thickness (t_{ox}) is 1000 Å, P-base dopant density (N_A) is 1×10^{17} cm^{-3}, gate doping is $N_D = 1 \times 10^{20}$cm^{-3}. The concentration of positively charged ions at the oxide–silicon interface is 1×10^{10} cm^{-2}. (b) If N_A is raised to 3×10^{17} cm^{-3}, what is the change in threshold voltage? (c) What is the impact of a higher oxide charge $= 5 \times 10^{11}$ cm^{-2}?

(a) The IGBT unit cell for this problem is shown in Fig. E3.2.1 Here, $q\phi_{\text{pSi-monoSi}} = (kT/q)\ln(N_{\text{DpSi}}N_A/n_i^2) - E_g = 0.0259 \times \ln\{1 \times 10^{20} \times 1 \times 10^{17}/(1.45 \times 10^{10})^2\} - 1.12 = -0.1254$ V. $(2kT/q)\ln(N_A/n_i) = (2 \times 0.0259) \times \ln(1 \times 10^{17}/1.45 \times 10^{10}) = 0.81567$ V. $C_{ox} = \varepsilon_0\varepsilon_{ox}/t_{ox} = 8.854 \times 10^{-14} \times 3.9/(1000 \times 10^{-8}) = 3.453 \times 10^{-8}$ F. $\surd\{4\varepsilon_0\varepsilon_s(kT)N_A \ln(N_A/n_i)\}/C_{ox} = \surd[4 \times 8.854 \times 10^{-14} \times 11.9 \times 1.38 \times 10^{-23} \times 300 \times 1 \times 10^{17} \times \ln\{1 \times 10^{17}/(1.45 \times 10^{10})]/3.453 \times 10^{-8} = 4.8$ V. This term of the threshold voltage equation is assigned a negative sign because the charge density in the depletion region is negative due to the negative acceptor ionic charges in the P-type material. So, this term $= -4.8$ V. $Q_{SS}/C_{ox} = 1 \times 10^{10} \times 1.6 \times 10^{-19}/3.453 \times 10^{-8} = 2.896 \times 10^{17} = 0.0463$ V. $\therefore V_{Th} = (kT/q)\ln(N_{\text{DpSi}}N_A/n_i^2) - E_g + (2kT/q)\ln(N_A/n_i) - \surd\{4\varepsilon_0\varepsilon_s kTN_A \ln(N_A/n_i)\}/C_{ox} - Q_{SS}/C_{ox} = -0.1254 + 0.81567 + 4.8 - 0.0463$ V $= 5.444$ V $= 5.4$ V.

(b) The second and third terms in the threshold voltage expression are altered on changing the P-base doping. The second term = 0.872578 V and the third term = 8.5995 V. Therefore, $V_{Th} = -0.1254 + 0.872578 + 8.5995 - 0.0463 = 9.3$ V, and the increase in threshold voltage $= 9.3 - 5.4 = 3.9$ V.

(c) The last term of threshold voltage expression becomes $= -5 \times 10^{11}$ cm$^{-2} \times 1.6 \times 10^{-19}/3.453 \times 10^{-8} = 2.3168$ V. Hence, $V_{Th} = -0.1254 + 0.81567 + 4.8 - 2.3168 = 3.17$ V. So, threshold voltage decreases by $5.4 - 3.17 = 2.23$ V.

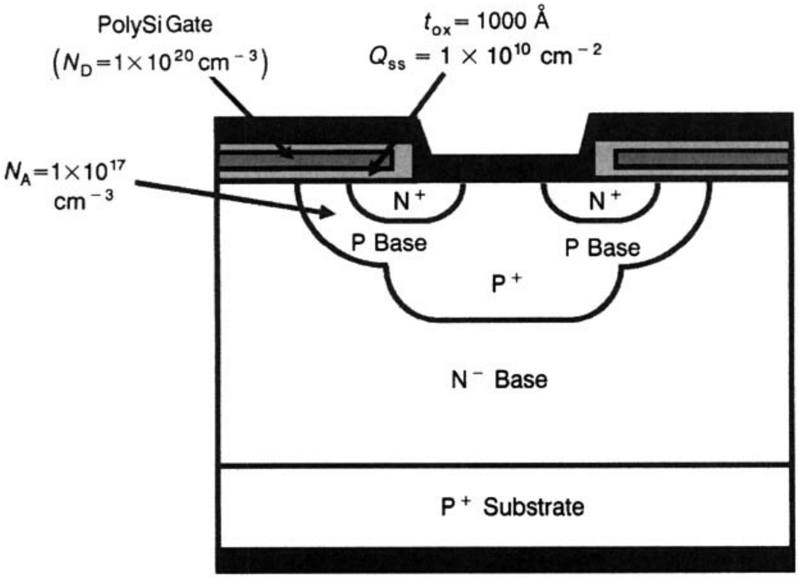

PolySi Gate
($N_D = 1 \times 10^{20}$ cm^{-3})
$t_{ox} = 1000$ Å
$Q_{ss} = 1 \times 10^{10}$ cm^{-2}
$N_A = 1 \times 10^{17}$ cm^{-3}
N+
N+
P Base
P Base
P+
N− Base
P+ Substrate

Figure E3.2.1 IGBT unit cell structure and physical parameters for Example 3.2(a). As compared to Example 3.2(a), in Example 3.2(b), $N_A = 3 \times 10^{17}$ cm^{-3}; and in Example 3.2(c) $Q_{SS} = 5 \times 10^{11}$ cm^{-2}.

3.3 CURRENT–VOLTAGE CHARACTERISTICS OF MOSFET; TRANSCONDUCTANCE AND DRAIN RESISTANCE

To derive the terminal current voltage ($i_{DS} - v_{DS}$) characteristics of the MOSFET, following simplifying assumptions are made: (i) The gate structure is an ideal MOS structure, as defined in Section 3.1 above. (ii) Carrier mobility in the inversion layer is constant, independent of the electric field intensity. (iii) Doping of the channel region is uniform. (iv) Carrier transport occurs only by the drift mechanism. (v) Reverse leakage current is very small and may be ignored. (vi) The *gradual channel approximation* is valid; that is, the transverse electric field produced by the gate voltage is much larger than the longitudinal electric field. Then under these idealized conditions, an N-channel MOSFET sketched in Fig. 3.8 is considered. Excluding the reversal of voltage and current polarities, the P-channel MOSFET operation is identical to the N-channel device. Let us focus our attention on an incremental length of the channel dy. Let the channel width (perpendicular to the plane of the diagram) be Z. For small values of drain-source voltage v_{DS}, the charge per unit area $Q(y)$ in the channel at a distance y from the source is given by [7]

$$Q(y) = C_{ox}\{v_{GS} - v(y) - V_{Th}\} \tag{3.16}$$

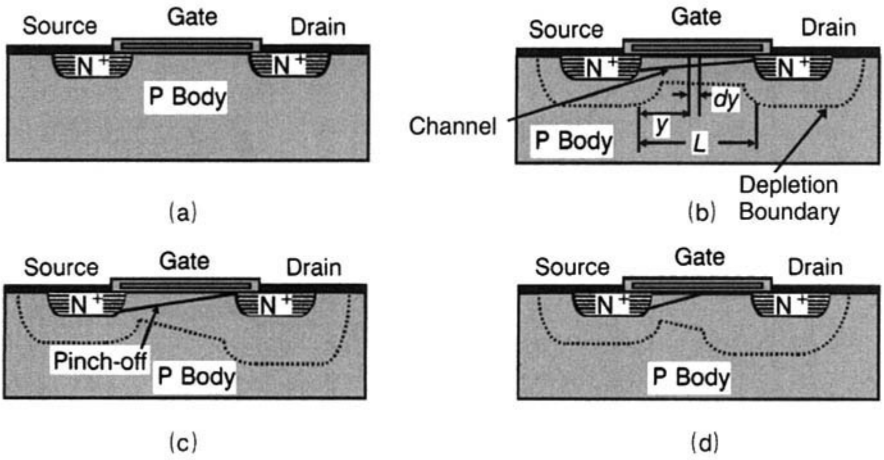


Figure 3.8 (a) Lateral MOSFET structure for studying the current–voltage characteristics; (b), (c), and (d) Channel charge distribution under drain current flow, for pinch-off condition, and beyond pinch-off point.

where C_{ox} is the capacitance per unit area. This equation is a statement of the definition of capacitance as charge stored per unit potential difference. It implies that the charge induced in the channel is determined by the potential drop = gate-source voltage − potential drop along the channel produced by the current flow − threshold voltage. Resistance of length dy of the channel is

$$dR = \frac{dy}{\mu_{\mathrm{ns}} Q(y) Z} \tag{3.17}$$

where μ_{ns} is the average electron mobility in the channel. This equation is obtained by noting that the conductivity of the channel is approximated by $\sigma(x) = qn(x)\mu_{\mathrm{ns}}(x)$, where q is the electronic charge and n is the electron concentration per unit volume. Then the channel conductance is $g = (Z/L)\int_0^{x_i} \sigma(x)\,dx$ where x_{i} denotes the point at which the intrinsic Fermi level intersects the quasi-Fermi level for electrons. For constant mobility μ_{ns}, this equation reduces to $g = (qZ\mu_{\mathrm{ns}}/L)\int_0^{x_i} n(x)\,dx = qZ\mu_{\mathrm{ns}}Q(y)/L$. So, the resistance of the element dy is $dR = dy/(gL) = dy/\{Z\mu_{ns}Q(y)\}$.

Applying Eq. (3.17), the voltage drop $dv(y)$ along the channel in the y direction with respect to the source is

$$dv(y) = i_{\mathrm{DS}}\, dR = \frac{i_{\mathrm{DS}}\, dy}{\mu_{\mathrm{ns}} Q(y) Z} \tag{3.18}$$

from which

$$i_{\mathrm{DS}}\, dy = Z\mu_{\mathrm{ns}} Q(y)\, dv(y) \tag{3.19}$$

Integration is performed along the channel from $y = 0$ to $y = L$:

$$
\begin{aligned}
\int_0^L i_{\mathrm{DS}}\, dy = i_{\mathrm{DS}} L &= \int_0^{v_{\mathrm{DS}}} Z\mu_{\mathrm{ns}} Q(y)\, dv(y) \\
&= \int_0^{v_{\mathrm{DS}}} Z\mu_{\mathrm{ns}} C_{\mathrm{ox}} \{v_{\mathrm{GS}} - v(y) - V_{\mathrm{Th}}\}\, dv(y) \\
&= Z\mu_{\mathrm{ns}} C_{\mathrm{ox}} \left\{ v_{\mathrm{GS}} \times [v(y)]_0^{V_{\mathrm{DS}}} - \left[\frac{v^2(y)}{2}\right]_0^{V_{\mathrm{DS}}} - V_{\mathrm{Th}} \times [v(y)]_0^{V_{\mathrm{DS}}} \right\} \\
&= Z\mu_{\mathrm{ns}} C_{\mathrm{ox}} \frac{2 v_{\mathrm{GS}} v_{\mathrm{DS}} - v_{\mathrm{DS}}{}^2 - 2 V_{\mathrm{Th}} v_{\mathrm{DS}}}{2} \qquad (3.20)
\end{aligned}
$$

yielding

$$
i_{\mathrm{DS}} = \frac{Z\mu_{\mathrm{ns}} C_{\mathrm{ox}}}{2L} \{2(v_{\mathrm{GS}} - V_{\mathrm{Th}}) v_{\mathrm{DS}} - v_{\mathrm{DS}}{}^2\} \qquad (3.21)
$$

The factor $(Z/2L)\mu_{\mathrm{ns}} C_{\mathrm{ox}} = K$ depends on the electrophysical parameters of MOS structure and its geometry. It is called the *specific transconductance* or *gain factor* of the transistor.

For small values of the drain–source voltage V_{DS}, the second term in Eq. (3.21) is negligibly small, yielding

$$
i_{\mathrm{DS}} \cong \frac{Z}{L} \mu_{\mathrm{ns}} C_{\mathrm{ox}} (v_{\mathrm{GS}} - V_{\mathrm{Th}}) v_{\mathrm{DS}} \qquad (3.22)
$$

from which the channel resistance r_{ch} is obtained as

$$
r_{\mathrm{ch}} = \frac{v_{\mathrm{DS}}}{i_{\mathrm{DS}}} = \frac{v_{\mathrm{DS}}}{\frac{Z}{L} \mu_{\mathrm{ns}} C_{\mathrm{ox}} (v_{\mathrm{GS}} - V_{\mathrm{Th}}) v_{\mathrm{DS}}} = \frac{L}{Z\mu_{\mathrm{ns}} C_{\mathrm{ox}} (v_{\mathrm{GS}} - V_{\mathrm{Th}})} \qquad (3.23)
$$

Amplifying properties of the MOSFET are described in terms of the transconductance parameter defined as

$$
g_{\mathrm{m}} = \left. \frac{\partial i_{\mathrm{DS}}}{\partial v_{\mathrm{GS}}} \right|_{v_{\mathrm{DS}} = \text{constant}} \qquad (3.24)
$$

In the linear region, using Eq. (3.22), the transconductance is given by

$$g_{\mathrm{ml}} = \left.\frac{\partial i_{\mathrm{DS}}}{\partial_{\mathrm{GS}}}\right|_{v_{\mathrm{DS}}} = \frac{d}{dv_{\mathrm{GS}}}\left\{\frac{Z}{L}\mu_{\mathrm{ns}}C_{\mathrm{ox}}(v_{\mathrm{GS}} - V_{\mathrm{Th}})v_{\mathrm{DS}}\right\}$$

$$= \frac{Z}{L}\mu_{\mathrm{ns}}C_{\mathrm{ox}}(1-0)v_{\mathrm{DS}} = \frac{Z}{L}\mu_{\mathrm{ns}}C_{\mathrm{ox}}v_{\mathrm{DS}} \tag{3.25}$$

At high v_{DS} values, the second term in Eq.(3.21) becomes larger. The reason is the reduction in channel inversion layer charge near the drain. When $v_{\mathrm{DS}} = v_{\mathrm{GS}} - V_{\mathrm{Th}}$, the channel charge at the drain end becomes zero. In this condition, known as *channel pinch-off*, the channel resistance $r_{\mathrm{ch}} = \infty$ and the drain–source current i_{DS} saturates (see Figs. 3.8c and 3.8d). Putting $v_{\mathrm{DS}} = v_{\mathrm{GS}} - V_{\mathrm{Th}}$ in Eq. (3.21), the saturated drain–source current $i_{\mathrm{DS(sat)}}$ is

$$i_{\mathrm{DS(sat)}} = i_{\mathrm{DS}} = \frac{Z\mu_{\mathrm{ns}}C_{\mathrm{ox}}}{2L}\left\{2(v_{\mathrm{GS}} - V_{\mathrm{Th}})(v_{\mathrm{GS}} - V_{\mathrm{Th}}) - (v_{\mathrm{GS}} - V_{\mathrm{Th}})^2\right\}$$

$$= \frac{Z}{2L}\mu_{\mathrm{ns}}C_{\mathrm{ox}}(v_{\mathrm{GS}} - V_{\mathrm{Th}})^2 \tag{3.26}$$

It represents the maximum current supported by the channel. Practical value of $i_{\mathrm{DS(Sat)}}$ is smaller than that given by Eq. (3.26) because of the reduction of μ_{ns} with longitudinal electric field. In the saturation region, the transconductance becomes

$$g_{\mathrm{ms}} = \left.\frac{\partial i_{\mathrm{DS}}}{\partial v_{\mathrm{GS}}}\right|_{v_{\mathrm{DS}}} = \frac{d}{dv_{\mathrm{GS}}}\left\{\frac{Z}{2L}\mu_{\mathrm{ns}}C_{\mathrm{ox}}(v_{\mathrm{GS}} - v_{\mathrm{Th}})^2\right\}$$

$$= \frac{Z}{2L}\mu_{\mathrm{ns}}C_{\mathrm{ox}} \times 2(v_{\mathrm{GS}} - v_{\mathrm{Th}})(1-0)$$

$$= \frac{Z}{L}\mu_{\mathrm{ns}}C_{\mathrm{ox}}(v_{\mathrm{GS}} - v_{\mathrm{Th}}) \tag{3.27}$$

In actual practice, the drain–source current i_{DS} slightly increases for $v_{\mathrm{DS}} \geq v_{\mathrm{GS}} - V_{\mathrm{Th}}$ due to decrease in effective channel length with increasing v_{DS} values, resulting in a finite drain–source output resistance defined as

$$r_{\mathrm{ds}} = \left.\frac{\partial v_{\mathrm{DS}}}{\partial i_{\mathrm{DS}}}\right|_{v_{\mathrm{GS}} = \text{constant}} \tag{3.28}$$

Voltage amplification produced by a MOSFET is

$$A_{\mathrm{V}} = \left.\frac{\partial v_{\mathrm{DS}}}{\partial v_{\mathrm{GS}}}\right|_{v_{\mathrm{GS}} = \mathrm{constant}} \tag{3.29}$$

A_{v}, g_{m} and r_{ds} are related as

$$A_{\mathrm{V}} = \left.\frac{\partial v_{\mathrm{DS}}}{\partial v_{\mathrm{GS}}}\right|_{v_{\mathrm{GS}}=\mathrm{constant}} = \left.\frac{\partial i_{\mathrm{DS}}}{\partial v_{\mathrm{GS}}}\right|_{v_{\mathrm{DS}}=\mathrm{constant}} \times \left.\frac{\partial v_{\mathrm{DS}}}{\partial i_{\mathrm{DS}}}\right|_{v_{\mathrm{GS}}=\mathrm{constant}} = g_{\mathrm{m}} r_{\mathrm{ds}} \tag{3.30}$$

3.4 ON-RESISTANCE MODEL OF DMOSFET AND UMOSFET

3.4.1 DMOSFET Model

The ON resistance of a power DMOSFET is a series combination of several resistance elements shown in Fig. 3.9. The specific ON resistance of a DMOSFET is expressed as the sum [8–11]

$$R_{\mathrm{on}} = R_{\mathrm{N+}} + R_{\mathrm{Ch}} + R_{\mathrm{A}} + R_{\mathrm{J}} + R_{\mathrm{D}} + R_{\mathrm{S}} \tag{3.31}$$

which contains the following components:

(i) $R_{\mathrm{N+}}$ is the resistance of the N^+ source diffusion, which is generally negligible.

(ii) R_{Ch} is the channel resistance given by

$$R_{\mathrm{Ch}} = \frac{L}{Z\mu_{\mathrm{ns}}C_{\mathrm{ox}}(V_{\mathrm{GS}} - V_{\mathrm{Th}})} \tag{3.32}$$

For the linear cell geometry, the channel resistance per square centimeter is (Appendix 3.4)

$$R_{\mathrm{Ch,sp}} = \frac{L(L_{\mathrm{G}} + 2m)}{2\mu_{\mathrm{ns}}C_{\mathrm{ox}}(V_{\mathrm{GS}} - V_{\mathrm{Th}})} \tag{3.33}$$

where $(L_{\mathrm{G}} + 2m)$ is the cell pitch.

(iii) R_{A} is the accumulation layer resistance accounting for the current spreading from the channel into the JFET region. Its value per square

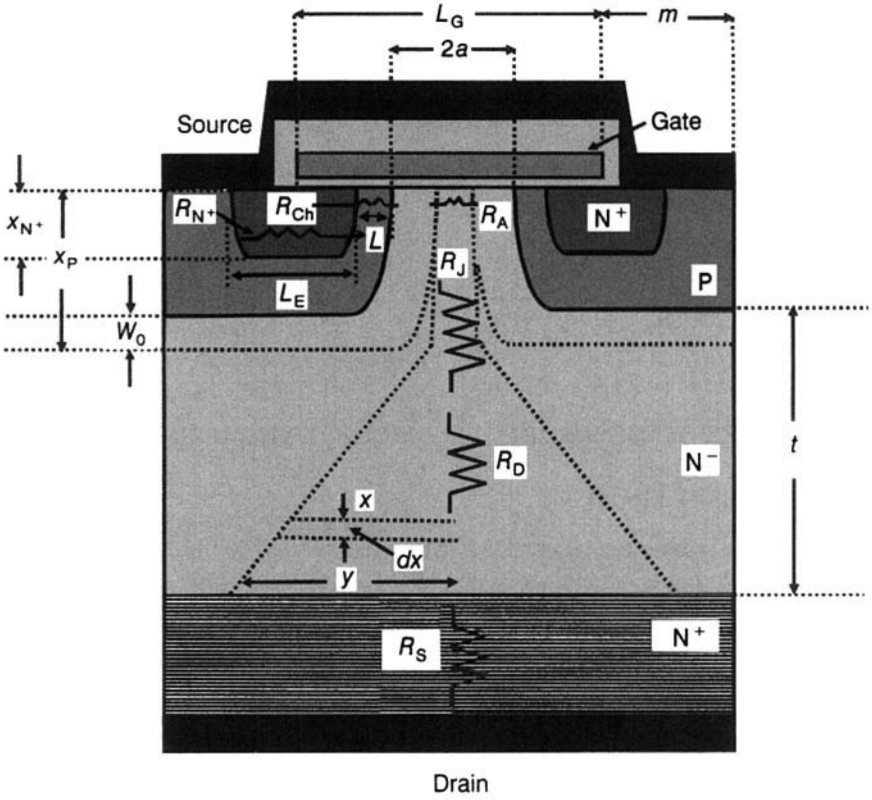


Figure 3.9 Illustrating the components of ON resistance of a DMOSFET.

centimeter for the linear cell structure is (Appendix 3.4)

$$R_{A,sp} = \frac{0.6(L_G - 2x_p)(L_G + 2m)}{2\mu_{nA} C_{ox}(V_{GS} - V_{Th})} \tag{3.34}$$

(iv) R_J is the JFET region resistance. It is the resistance of the drift region between P-base diffusions acting as the gate regions of the JFET (see Section 4.5). Resistance of the JFET region is readily derived by neglecting the influence of voltage drop along the vertical direction on the depletion region. Assuming uniform current flow from the accumulation layer into the JFET region, resistance of this region is simply that of a semiconductor with a cross-sectional area given by

$$A = aZ = \left(\frac{L_G}{2} - x_p\right)Z$$

where Z is the cell width perpendicular to the cross section. Then (Appendix 3.4)

$$R_{\mathrm{J,sp}} = \frac{\rho_{\mathrm{D}}(L_{\mathrm{G}} + 2m)(x_{\mathrm{p}} + W_0)}{(L_{\mathrm{G}} - 2x_{\mathrm{p}} - 2W_0)} \tag{3.35}$$

where ρ_{D} is the resistivity of the JFET region and W_0 is the depletion layer width extending under the gate.

(v) R_{D} is the drift region resistance. For current spreading from a cross-section of $a = L_{\mathrm{G}} - 2x_{\mathrm{P}}$ at 45° angle, the cross section for current flow at a depth x below the P-base region is $y = (a + x)Z$. Spreading resistance of the drift region is

$$R_{\mathrm{D}} = \int_0^t \frac{\rho_{\mathrm{D}}}{(a+x)Z}\,dx = \frac{\rho_{\mathrm{D}}}{Z}\int_0^t \frac{dx}{(a+x)}$$
$$= \frac{\rho_{\mathrm{D}}}{Z}[\ln(a+x)]_0^t = \frac{\rho_{\mathrm{D}}}{Z}[\ln(a+t) - \ln a] = \frac{\rho_{\mathrm{D}}}{Z}\ln\left(\frac{a+t}{a}\right)$$

where t is the thickness of the drift layer, and the formula $\int dx/(a+x) = \ln(a+x)$, has been used. The specific resistance is (Appendix 3.4)

$$R_{\mathrm{D,sp}} = \frac{\rho_{\mathrm{D}}(L_{\mathrm{G}} + 2m)}{2}\ln\left(\frac{a+t}{a}\right) \tag{3.36}$$

(vi) R_{S} is the substrate resistance:

$$R_{\mathrm{S,sp}} = \rho_{\mathrm{S}} t_{\mathrm{S}} \tag{3.37}$$

where ρ_{S} and t_{S} denote the resistivity and thickness of the substrate, respectively.

Design optimization of a power DMOSFET structure involves extensive simulations of polySi gate length to obtain the minimum specific ON resistance. When the polySi gate length is large, the contributions of JFET and drift regions to ON resistance decrease. At the same time, the accumulation layer resistance increases because of the longer path length traversed by the carriers along the surface. Also, the channel resistance decreases due to larger cell pitch resulting in a decrease of channel density. Reversal of the pattern of variation of these parameters occurs when the polySi gate length is made small. Thus the device designer seeks the minimum value of specific ON resistance obtainable with respect to gate length.

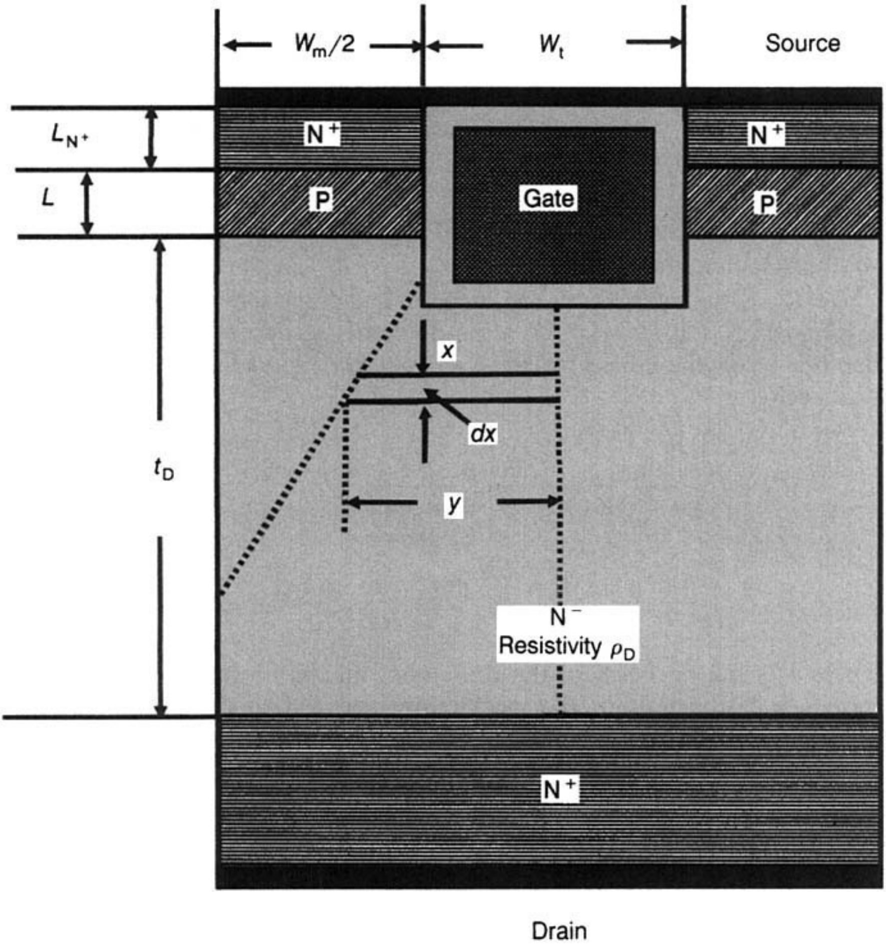

$W_m/2$
W_t
Source
L_{N^+}
L
N+
P
Gate
N+
P
x
dx
t_D
y
N−
Resistivity ρ_D
N+
Drain

Figure 3.10 ON-resistance components of UMOSFET.

3.4.2 UMOSFET Model

In the UMOSFET configuration (Fig. 3.10), there is no JFET region. The current spreads out from an accumulation layer produced on the surface of the U groove into the drift region. The main components of ON resistance of UMOSFET cell are the channel resistance and drift region resistance [12–14]. If W_m is the mesa width and W_t is the trench width, the channel resistance of UMOSFET structure per unit area is

$$R_{\mathrm{Ch,sp}} = \frac{L(W_m + W_t)}{2\mu_{ns}C_{ox}(V_{GS} - V_{Th})} \tag{3.38}$$

This equation follows straightforward from Appendix 3.4, Eq. (A3.4.3) for the specific channel resistance of the linear DMOSFET cell, by noting that the area of the UMOSFET unit cell $= A = (W_\mathrm{m} + W_\mathrm{t})Z$ instead of $(L_\mathrm{G} + 2m)Z$ for the DMOSFET cell. Absence of JFET region in UMOSFET allows the fabrication of narrow mesa and trench regions. It is advantageous to decrease the mesa and trench widths as much as practically feasible. By fine-resolution lithographic techniques, the cell size can be reduced below 6 μm. Consequently, much higher channel densities (channel width per unit active cell area) are achieved with UMOSFET design as compared to DMOSFET or VMOSFET structures. The overall impact is that the contribution of channel resistance to the ON resistance falls. The reason why this is significant is that the ON resistance of the UMOSFET is lowered much below that of the remaining categories of MOSFETs, making the conduction power loss very small. Figure 3.11 pictorially displays the comparative analysis of the UMOSFET and DMOSFET classes.

For the UMOSFET structure, the spreading resistance of the drift region is written as [15] (see Appendix 3.5)

$$R_\mathrm{D} = \rho_\mathrm{D}\left\{0.5(W_\mathrm{m} + W_\mathrm{t})\ln\left(1 + \frac{W_\mathrm{m}}{W_\mathrm{t}}\right) + (t_\mathrm{D} - 0.5W_\mathrm{m})\right\} \tag{3.39}$$

The first term describes the drift region segment where current spreads at 45°. The second term is concerned with the drift region where cross-sectional area equals the cell area. Current overlapping takes place even for low-voltage designs because of the very small mesa widths used. Then the spreading resistance of the drift region approaches the ideal specific ON resistance. Hence, the UMOSFET structure allows the fabrication of devices with specific ON resistance close to the ideal value.

Thus, it is evident that that as the device designer progresses toward smaller mesa and trench widths of the UMOSFET, the channel resistance is reduced and the channel density increases. Additionally, the spreading resistance of drift region approaches the ideal limiting value of specific ON resistance. Beyond these limits, further lowering of specific ON resistance is obtained by digging the trench up to the N^+ substrate. Then the drain current flow is divided between the drift region and the trench sidewalls via the accumulation layer. The breakdown voltage of such MOSFET structures is very low (~ 25 V) because the drain voltage is entirely supported by the gate oxide. By using a dual thickness gate oxide structure with smaller thickness up to the P-base bottom and larger thickness thereafter, the breakdown voltage is increased. At the same time, the accumulation layer resistance in the thicker gate oxide region is raised.

Example 3.3 Calculate the specific ON resistance of an N-channel power DMOSFET having linear cell geometry, operating with a gate bias $V_\mathrm{GS} = 10$ V and $V_\mathrm{DS} =$ 3 V, given the following structural parameters of the device: polySi gate length

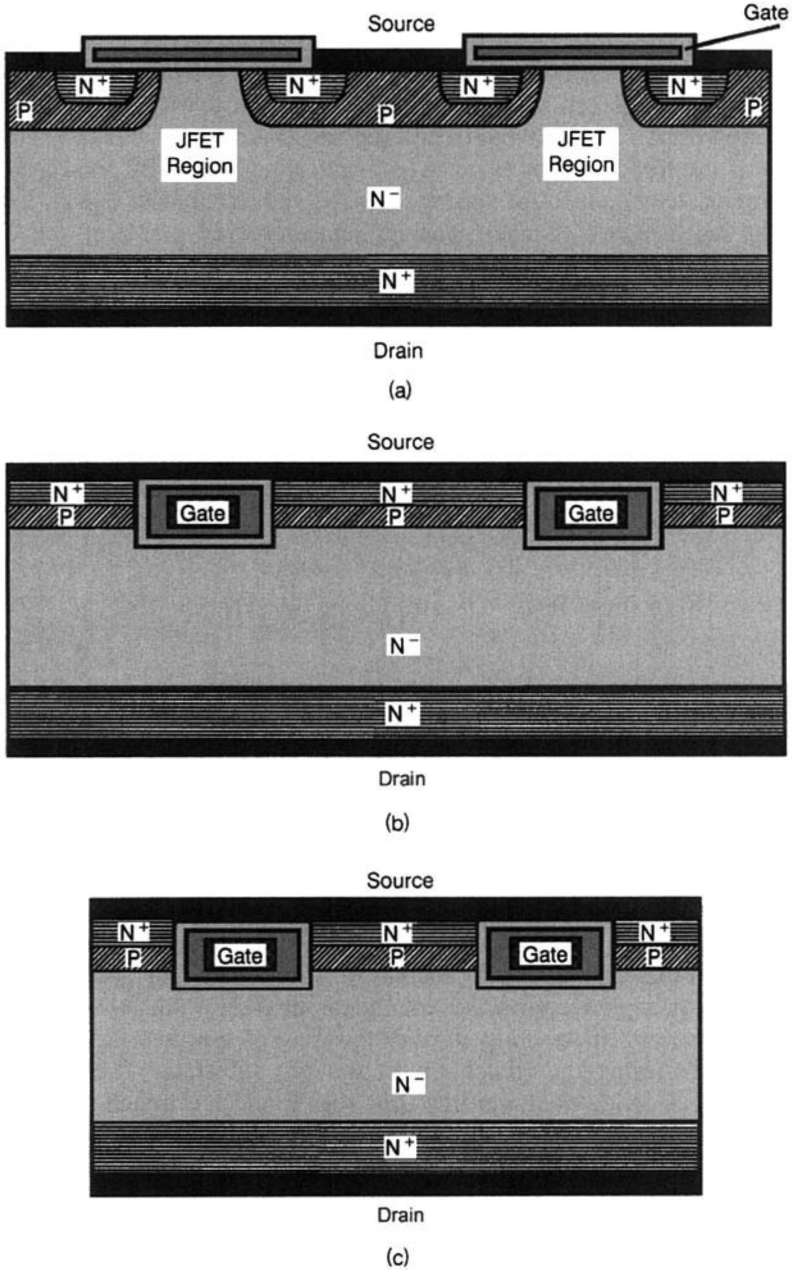

Source
Gate
N+
N+
N+
N+
P
P
P
JFET Region
JFET Region
N−
N+
Drain
(a)
Source
N+
P
Gate
N+
P
Gate
N+
P
N−
N+
Drain
(b)
Source
N+
P
Gate
N+
P
Gate
N+
P
N−
N+
Drain
(c)

Figure 3.11 Comparison of DMOSFET and UMOSFET structures. (a) DMOSFET cross section: Constriction of the JFET region intervening neighboring P-base cells will increase the JFET resistance. (b) UMOSFET cross section: Absence of JFET region allows reduction of cell size as shown in part (c). (c) Cell size reduction and enhancement of current density capability provided by UMOSFET structure.

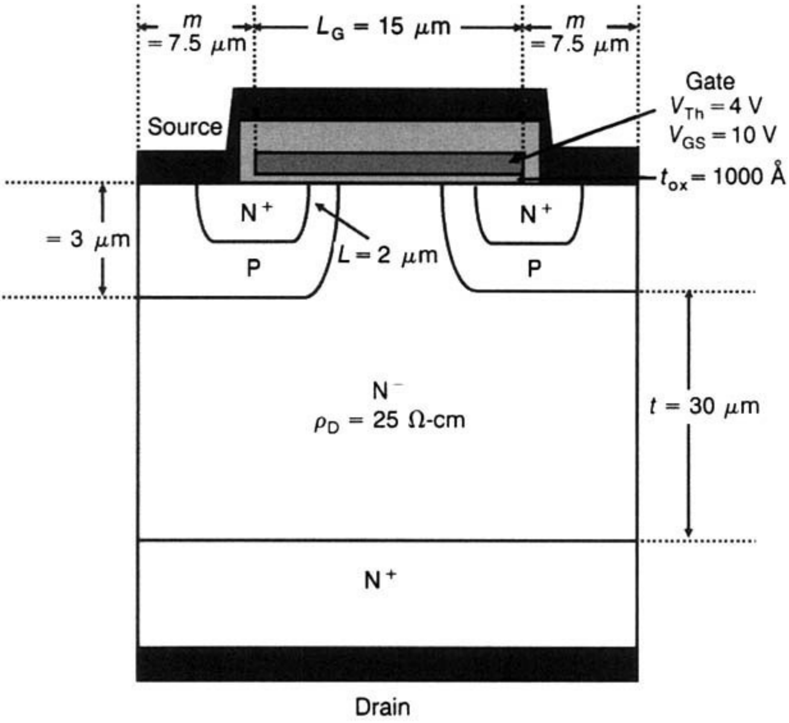


Figure E3.3.1 DMOSFET structure for Example 3.3.

$(L_G) = 15\ \mu$m, cell pitch $(L_G + 2m) = 30\ \mu$m, channel length $(L_{Ch}) = 2\ \mu$m, gate oxide thickness $(t_{ox}) = 1000$ Å, threshold voltage $(V_{Th}) = 4$ V, P-base depth $(x_P) =$ 3 μm, resistivity of N$^-$-drift region $(\rho_D) = 25\ \Omega$-cm, its thickness $(t) = 30\ \mu$m, and the surface electron mobility value = 500 cm^2/V-sec. What percentage of the total ON resistance is the JFET resistance?

Figure E3.3.1 shows the DMOSFET unit cell. Since $C_{ox} = 8.854 \times 10^{-14} \times 11.9/1000 \times 10^{-8} = 1.0536 \times 10^{-7}$ F/cm^2, the channel specific resistance is $R_{Ch,SP} = 2 \times 10^{-4} \times 30 \times 10^{-4}/(2 \times 500 \times 1.0536 \times 10^{-7} \times 6) = 9.49 \times 10^{-4}$ Ω-cm^2. $R_{ACC,SP} = \{\{0.6 \times (15 - 2) \times 10^{-4} \times (30 \times 10^{-4})\}/(2 \times 500 \times 1.0536 \times 10^{-7} \times 6) = 3.7 \times 10^{-3}$ Ω-cm^2. For $\rho = 25$ Ω-cm, dopant density $N_D = 1/(\rho e \mu) = 1/(25 \times 1.6 \times 10^{-19} \times 1350) = 1.85 \times 10^{14}$ cm^{-3}. At $V_{DS} = 3$ V, $W_0 = \surd\{2\varepsilon_0\varepsilon_s V/(qN_A)\} = \surd\{2 \times 8.854 \times 10^{-14} \times 3.9 \times 3/(1.6 \times 10^{-19} \times 1.85 \times 10^{14}\} = 2.646 \times 10^{-4}$ cm. $\therefore$JFET resistance $R_{J,SP} = \{25 \times 30 \times 10^{-4} \times (3 + 2.646) \times 10^{-4}\}/\{(15 - 2 \times 3 - 2 \times 2.646) \times 10^{-4}\} = 0.114199$ Ω-cm^2. Drift region resistance $R_{D,SP} = (25 \times 30 \times 10^{-4}/2) \ln\{(9 + 30)/9\} = 0.0549876$ Ω-cm^2. $\therefore R_{ON,SP} = 9.49 \times 10^{-4} + 3.7 \times 10^{-3} + 0.114199 + 0.0549876 = 0.1738366$ Ω-cm^2. Now, $R_{J,SP}/R_{ON,SP} \times 100\%$ $= 0.114199 \times 100\%/0.1738366 = 65.7\%$.

Example 3.4 Repeat the above calculations for UMOSFET device with identical parameters, if mesa and trench width $(W_m + W_t) = 6\ \mu$m.

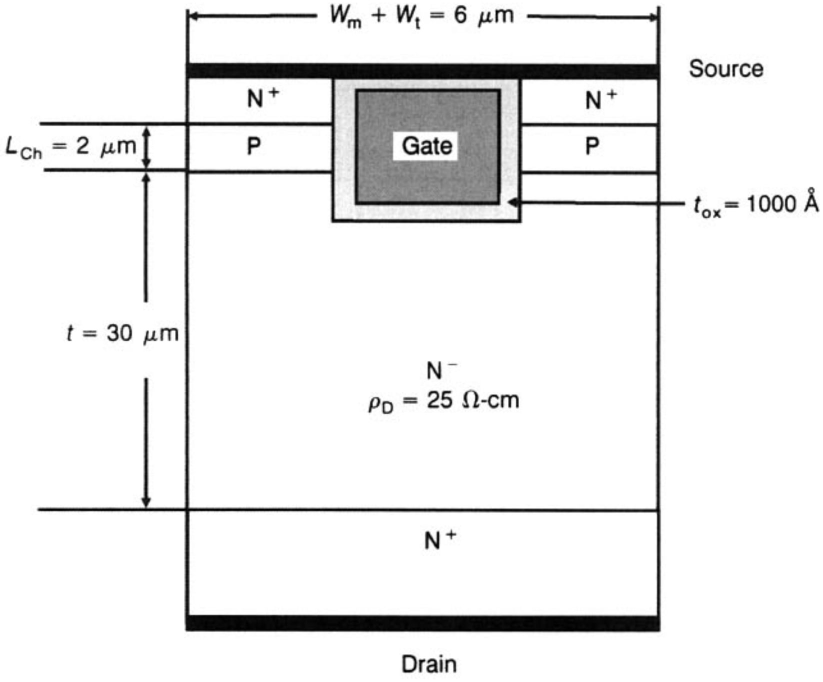


Figure E3.4.1 UMOSFET cell for Example 3.4.

Figure E3.4.1 depicts the UMOSFET unit cell. Here, $R_{\mathrm{Ch,SP}} = 2 \times 10^{-4} \times 6 \times 10^{-4}/(2 \times 500 \times 1.0536 \times 10^{-7} \times 6) = 1.898 \times 10^{-4}$ Ω-cm². $R_{\mathrm{D,SP}} = 25\{(6 \times 10^{-4}/2)\ln(6/3) + (30 - 3/2) \times 10^{-4}\} = 0.07645$ Ω-cm². $R_{\mathrm{ON,SP}} = 0.07664$ Ω-cm².

3.5 MOSFET EQUIVALENT CIRCUIT AND SWITCHING TIMES

The electrical equivalent circuit of the MOSFET is shown in Fig. 3.12. This circuit reflects the physical structure of the MOSFET. Each circuit component has a specific physical meaning. The current generator $g_m v_{GS}$ in parallel with the drain-source output resistance R_O describes the amplifying properties of the MOSFET. The body resistances of the heavily doped source and drain regions, R_S and R_D respectively, are generally negligible. The channel resistance is r_{ch} (not shown in the figure) and the body resistance of the substrate is R_{Sub}. C_{GS} is the gate–source capacitance due to overlap between gate metallization and source region. The charge developed over C_{GS} determines the channel conductivity. Components of C_{GS} are (i) C_{N+} due to overlap of gate electrode over N$^+$ source, (ii) C_P arising from overlapping of

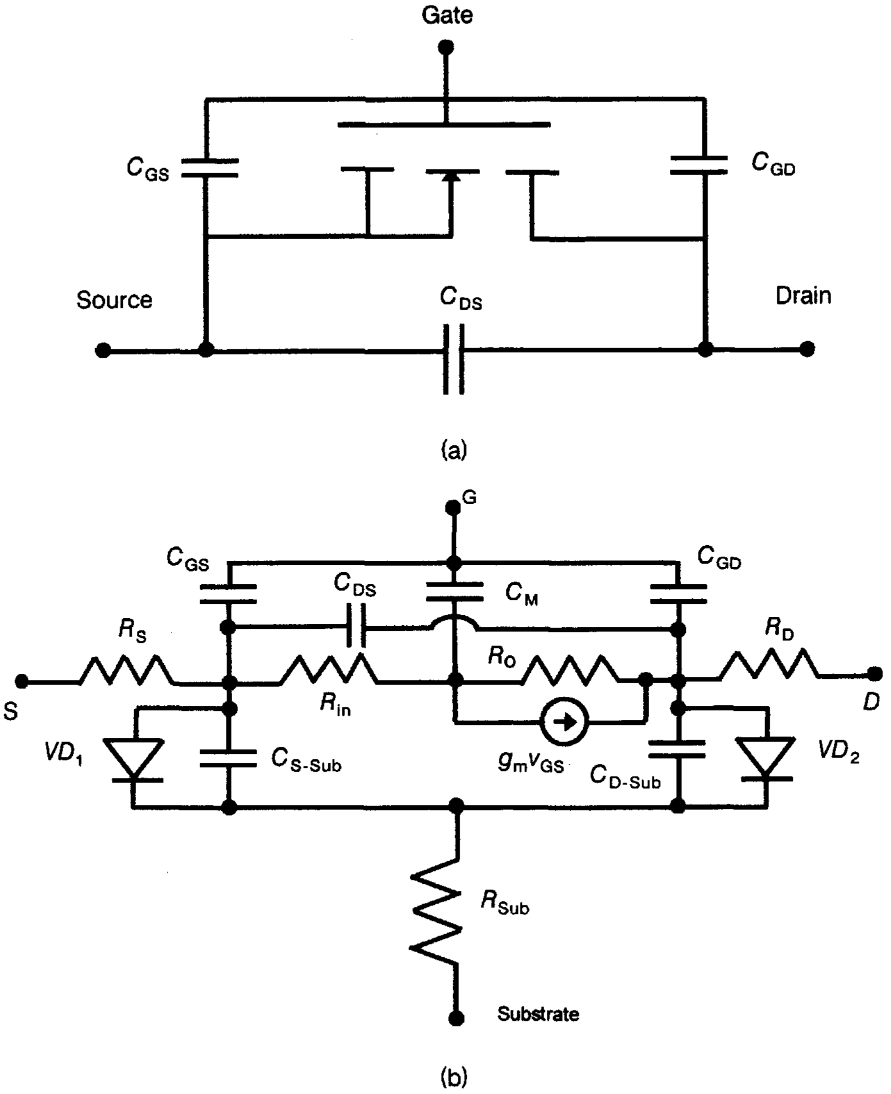

Gate
C_GS
C_GD
Source
C_DS
Drain
(a)
G
C_GS
C_DS
C_M
C_GD
R_S
R_D
R_O
R_in
S
D
VD_1
C_S-Sub
g_mV_GS
C_D-Sub
VD_2
R_Sub
Substrate
(b)

Figure 3.12 (a) MOSFET interelectrode capacitances and (b) electrical equivalent circuit of a MOSFET. **Notation:** C_{GS} = gate–source capacitance, C_{GD} = gate–drain capacitance, C_{DS} = drain–source capacitance, C_M = Miller capacitance, $C_{S\text{-Sub}}$, $C_{D\text{-sub}}$ = capacitances arising from the overlap of the junction between the doped region and the substrate; R_s, R_0 = body resistances of the highly doped source and drain regions; R_{Sub} = body resistance of the substrate, R_{in} = input resistance, R_0 = output resistance; V_{D1}, V_{D2} = source–substrate and drain–substrate junctions under reverse bias; g_mV_{GS} = current generator.

gate electrode on P base, and (iii) C_O produced by the source metal running over the gate electrode. C_{GD} is the gate–drain capacitance created by the overlapping of gate metallization with drain region. The input capacitance of the gate terminal with source and drain connected together is

$$C_{ISS} = C_{GS} + C_{GD} \tag{3.40}$$

Similarly, the output capacitance with the gate tied to source is

$$C_{OSS} = C_{DS} + C_{GD} \tag{3.41}$$

where C_{DS} is the drain-source capacitance. Capacitances $C_{S\text{-}Sub}$ and $C_{D\text{-}Sub}$ are produced by the overlapping of the junction between substrate and corresponding doped region. VD_1 and VD_2 represent the reverse–biased source–substrate and drain–substrate junctions.

C_{GD} is amplified by the Miller effect into an equivalent capacitance

$$C_M = (1 + g_{in}R_O)C_{GD} \tag{3.42}$$

where g_m denotes the transconductance and R_O is the output resistance. C_{GD} has a high value during the conduction state of the MOSFET when the surface of the drift region is in the accumulation condition but as V_{DS} increases, C_{GD} decreases. Amplification of this capacitance by Miller effect can severely reduce the frequency response. Restricting the overlapping of the gate over the drift region decreases C_{GD}, but the resultant high electric field at the gate edge lowers the cell breakdown voltage and increases the ON resistance, raising the resistance between the channel and the drift region because of the disappearance of accumulation layer over a portion of the surface between base regions. A shallow P diffusion at the interruption of the gate electrode helps in improving the cell breakdown voltage.

With reference to the waveforms of the switching control voltage v_{GS} and voltage across the switch v_{DS} (Fig. 3.13), four delay intervals are specified for a MOSFET device: (i) *Turn-on delay time* $t_{d(on)}$: Time interval reckoned from the instant v_{GS} has risen to 0.1th part of its final value to the instant v_{DS} has fallen by 0.1th part of its initial OFF-state value. (ii) *Rise time* t_r: Time interval during which v_{DS} falls from 0.9 to 0.1 fraction of its initial OFF-state value. (iii) *Turn-off delay time* $t_{d(off)}$: Time interval measured from the instant v_{GS} has fallen to 0.9 of its ON-state value to the instant v_{DS} has risen to 0.1 of its final OFF value. (iv) *Fall time* t_f: Time interval during which v_{DS} rises from 0.1 to 0.9 of its final OFF-state value.

In terms of these intervals, the turn-on time is

$$t_{ON} = t_{d(on)} + t_r \tag{3.43}$$

and the turn-off time is

$$t_{OFF} = t_{d(off)} + t_f \tag{3.44}$$

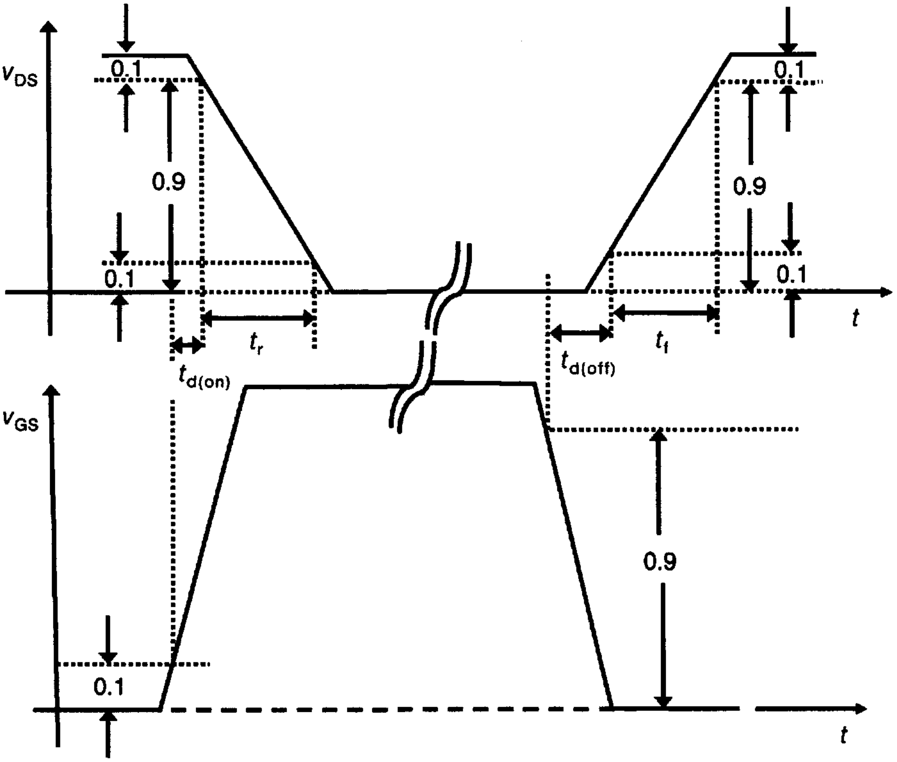


Figure 3.13 Switching times of a MOSFET showing both the output voltage v_{DS} and the input control voltage v_{GS}.

For clamped inductive load switching, if V_S is the supply voltage, V_F is the forward voltage drop across the MOSFET, R_G is the series resistance of the gate drive voltage supply, V_G is the gate drive voltage, C_{GD} is the gate-to-drain capacitance, and I_L is the steady-state current flowing through the load, the turn-on time of MOSFET is expressed as [16]

$$t_{on} = \frac{(V_S - V_F) R_G C_{GD}}{V_G - (V_T + I_L/g_m)} \tag{3.45}$$

and turn-off time is given by [16]

$$t_{off} = R_G (C_{GS} + C_{GD}) \ln\left(\frac{I_L}{g_m V_T} + 1 \right) \tag{3.46}$$

By reducing R_G and C_{GD}, dynamic power losses are lowered.

Example 3.5 In a power IGBT circuit, the constant current gate supply gives a charge of 10 nC to the gate electrode for a $\Delta v_{GE} = 5$ V. What is the input capacitance C_{iss} and how much energy is stored in this capacitance at $v_{GE} = 15$ V?

The gate current i_{GE} is given by

$$i_{GE} = C_{iss}\frac{dv_{GE}}{dt} \tag{E3.3.1}$$

If the gate drive (v_{GE} – charge Q) characteristic is linear, then we have

$$I_{GE} = C_{iss}\frac{\Delta v_{GE}}{\Delta t} \tag{E3.3.2}$$

yielding

$$C_{iss} = I_{GE}\frac{\Delta t}{\Delta v_{GE}} = \frac{\Delta Q}{\Delta v_{GE}} \tag{E3.3.3}$$

Thus $C_{iss} = 10 \times 10^{-9}/5 = 2 \times 10^{-9}$ F. Energy stored in this capacitance is

$$E_G = \tfrac{1}{2}C_{iss}V_{GE}{}^2 \tag{E3.3.4}$$

So, $E_G = \frac{1}{2} \times 2 \times 10^{-9} \times (15)^2 = 2.25 \times 10^{-7}$ J.

3.6 SAFE OPERATING AREA (SOA)

Figure 3.14 illustrates a typical safe operating area whose boundaries are prescribed by the maximum permitted drain–source current, the peak power dissipation, and the drain–source breakdown voltage. These limits are lower for continuous operation than for pulsed-mode operation. Furthermore, in the pulsed-mode operation, the limits are smaller for longer duration pulses. In MOSFETs, there are two kinds of second breakdown, namely, bipolar second breakdown and MOS second breakdown. These modes are differentiated in Table 3.2.

3.7 NEUTRON AND GAMMA-RAY DAMAGE EFFECTS

As MOSFETs are majority-carrier devices, the influence of neutron radiation on minority carrier lifetime does not impair device behavior. As already discussed in Section 2.7, gamma rays build up a trapped positive hole charge and interface fast states, producing changes in the threshold voltage, flatband voltage (defined in Example 3.1), and channel mobility. In NMOSFET, a positive gate bias is applied, so the trapped positive hole charge is repelled

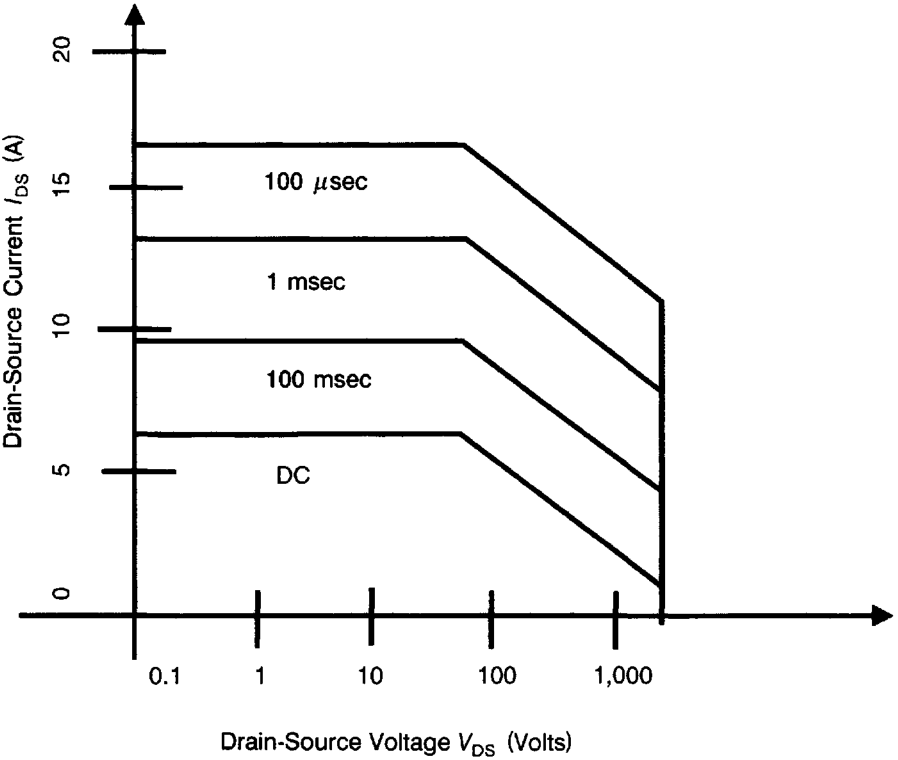


Figure 3.14 Typical safe operating area of a MOSFET.

toward the Si–SiO$_2$ interface where it is closer to the channel and causes more threshold voltage shift. In PMOSFET, the applied gate bias is negative; hence the trapped positive hole charge is attracted toward the metal gate electrode, away from the channel, producing smaller threshold voltage change. Consequently, the PMOS transistor is more radiation-resistant than the NMOS transistor. For both types of transistors, the threshold voltage shifts toward decreasing voltages. The PMOS transistor becomes harder to turn on, whereas the NMOS transistor becomes easier to turn on and could be made to turn on permanently. In the NMOS transistor, in which ionizing radiation has produced trapped hole charge in the field oxide (i.e., oxide other than the gate oxide), application of a positive gate bias can easily invert the Si beneath, producing leakage current. However in the PMOS transistor, the applied negative gate bias pulls the trapped hole charge toward the gate terminal, away from the channel, a less sensitive position for disrupting device operation.

Table 3.2 Mechanisms of Second Breakdown in MOSFETs

Serial No.	Bipolar Second Breakdown	MOS Second Breakdown
1	Originates from the presence of (N^+ source–P-base–N^--drift layer) parasitic bipolar transistor in the structure. On increasing the drain voltage to near breakdown, a large avalanche current flows laterally in the P-base in addition to the normal channel current. This large laterally flowing current in the P-base produces a potential drop in this region. When this potential drop exceeds the built-in potential of N^+ source–P-base junction, this junction is forward-biased and the N^+PN^- transistor is turned on. The P-base–N-drift layer breakdown voltage (BV_{CBO}) is reduced to BV_{CEO} and thus the parasitic transistor suffers from breakdown.	Arises from the influence of lateral voltage drop in the P-base region upon the channel current. The large electric field developed initiates avalanche multiplication of channel current, causing breakdown.
2	The drain voltage at which bipolar second breakdown occurs is given by $$V_{\mathrm{D,SB}} = \frac{BV}{\left\{1 + \left(\frac{qR_{\mathrm{B}}I_0}{kT}\right)\right\}^{1/n}}$$ where BV is the junction breakdown voltage, R_{B} is the base resistance, I_0 is the reverse saturation current, and n is a coefficient.	The drain voltage at which MOS second breakdown takes place is expressed as $$V_{\mathrm{D,SB}} = \frac{BV}{(1 + \gamma R_{\mathrm{B}})^{1/n}}$$ where γ is a body-bias coefficient = $\Delta I_{\mathrm{D}}/\Delta V_{\mathrm{B}}$.
3	As the base resistance increases, the second breakdown voltage decreases. So the incorporation of a deep P^+ diffused region in the DMOSFET cell increases the $V_{\mathrm{D,\,SB}}$ capability of the device.	The higher the R_{B} value, the lower the second breakdown voltage. Hence the deep P^+ diffusion helps improve the $V_{\mathrm{D,\,SB}}$ capability as in the bipolar case.

3.8 THERMAL BEHAVIOR OF MOSFET

Heating of a MOSFET has the combined effect of increasing the forward current by threshold voltage lowering and decreasing the same due to mobility reduction. The combined effect is that for a MOSFET operating near the threshold voltage, the output current increases with temperature while for a MOSFET working remote from it, the current decreases with temperature [17–19]. The mobility in a strongly inverted MOSFET channel is inversely proportional to the square of absolute temperature:

$$\mu(T) = \mu(300)\left(\frac{T}{300}\right)^{-2} \tag{3.47}$$

The variation of threshold voltage with temperature is expressed as [15-17]

$$\frac{dV_T}{dT} = \frac{d\psi_B}{dT}\left(\frac{1}{C_{ox}}\sqrt{\frac{\varepsilon_0\varepsilon_s N_A}{\psi_B}} + 2\right) \tag{3.48}$$

where (Appendix 3.6)

$$\frac{d\psi_B}{dT} = \frac{1}{T}\left\{\frac{E_g(T=0)}{2q} - |\psi_B(T)|\right\} \tag{3.49}$$

The above relationship follows from the thermal independence of work-function difference and fixed oxide charges.

Example 3.6 The P-base concentration (N_A) in an IGBT is 1×10^{17} cm^{-3} and the gate oxide thickness (t_{ox}) is 1000 Å. Calculate the threshold voltage (V_T) at −20°C, 27°C, 100°C, 200°C, and 300°C. Also determine the rate of change of threshold voltage with temperature ($\Delta V_T/\Delta T$), taking every two consecutive temperatures. It is given that the intrinsic carrier concentration is 1×10^8 cm^{-3}, 1.5×10^{10} cm^{-3}, 2×10^{12} cm^{-3}, 1×10^{14} cm^{-3}, and 2×10^{15} cm^{-3}, respectively, at these temperatures. The thermal voltages ($=kT/q$) are 0.0218 V, 0.0259 V, 0.0322 V, 0.0408 V, and 0.0494 V, respectively. The work function difference and oxide charge effects may be neglected.

From Eq. (3.12), we have $C_{ox} = 8.854 \times 10^{-14} \times 3.9/(1000 \times 10^{-8}) = 3.45306 \times 10^{-8}$ F/cm^2. Applying Eq. (3.14), $V_T = \sqrt{(4\varepsilon_0\varepsilon_s q N_A \psi_B)}/C_{ox} + 2\Psi_B$, and $\psi_B = (kT/q)\ln(N_A/n_i)$. The results are compiled in Table E3.6.1:

Table E3.6.1 Variation of Threshold Voltage with Temperature

Serial No.	Temperature (°C)	ψ_B (V)	V_T (V)	ΔV_T (mV)	$\Delta V_T/\Delta T$ (mV/°C)
1	−20	0.4522	5.96		
2	27	0.407	5.611	349	7.43
3	100	0.371	5.14	471	6.45
4	200	0.282	4.556	584	5.84
5	300	0.193	3.69	866	8.66

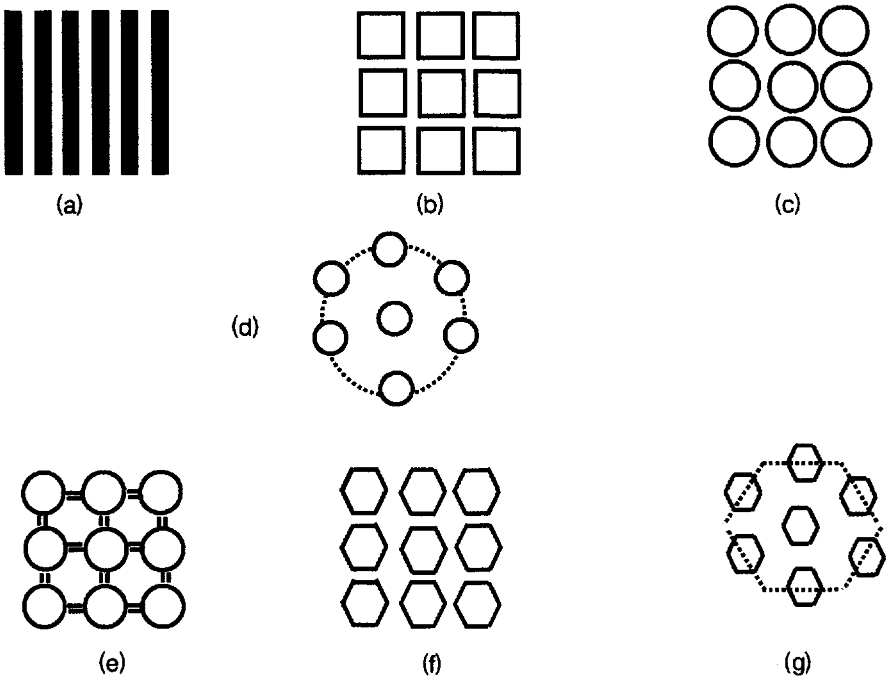


Figure 3.15 Different topological designs of DMOSFET unit cell: (a) Linear, (b) square cell in square array, (c) circular cell in square array, (d) circular cell in circular array, (e) atomic lattice layout (A-L-L), (f) hexagonal cell in square array, and (g) hexagonal cell in hexagonal array.

3.9 DMOSFET CELL WINDOWS AND TOPOLOGICAL DESIGNS

Cell windows commonly employed in power DMOSFETs include linear, square, circular, hexagonal, and atomic lattice layout (A-L-L) shapes. These cellular layouts are shown in Fig. 3.15. Assuming that the doping concentration (N_D) of the drift region is same for all geometries, their ON resistances are equal if the ratio (area of cell window/total area) is the same. But N_D needs to be adjusted to avoid electric field crowding in the cell. Then A-L-L is found to be superior to other geometries. It also gives a smaller drain overlap capacitance. This factor is desirable for high frequency operation.

3.10 SUMMARY AND TRENDS

The DMOSFET and UMOSFET devices are essential components of efficient and high-speed power electronic conversion and control systems for various industrial sectors. Technology growth in this vital power sector (up to > 300 V) rests on the discrete MOSFET foundation. Trench-gate structures have helped in cutting down the ON resistance substantially.

REVIEW EXERCISES

3.1 What do you understand by enhancement and depletion MOSFETs? Name the most commonly used type of MOSFET?

3.2 Why is the vertical structure preferred for power MOSFETs? Draw the cross-sectional diagrams of vertical and lateral MOSFETs, and describe their operation.

3.3 How is the integral body diode of a MOSFET used as a free-wheeling diode? Identify the parasitic bipolar transistor in a MOSFET and explain how it is responsible for device failure?

3.4 For the same blocking voltage, why does a MOSFET carry less forward current than a bipolar transistor? Why is a bipolar transistor slower than a MOSFET?

3.5 Does a MOSFET require substantial input current for DC operation? What happens at high frequencies?

3.6 What is meant by accumulation, depletion, and inversion of a semiconductor? Explain the terms strong, moderate, and weak inversion.

3.7 Distinguish the following: fixed oxide charge, trapped oxide charge, mobile oxide charge, and interface trap charge.

3.8 What do you understand by the flat-band voltage V_{FB} of a semiconductor? Write the formula for flat-band voltage and explain the symbols.

3.9 What simplifying assumptions are made in the derivation of the output current–voltage characteristics of a MOSFET. Derive the i_{DS}-v_{DS} relationship and show how it is modified for small values of v_{DS}. Define transconductance of a MOSFET.

3.10 Show on a diagram the internal resistances of a DMOSFET. Write the equations for the specific resistance due to the JFET and the drift regions.

3.11 Draw the electrical equivalent circuit of a MOSFET and explain the origin of the different circuit components.

3.12 Name and define the four principal delay times in MOSFET switching.

3.13 Draw the representative safe operating area of a MOSFET and explain how its boundaries are defined.

3.14 Explain why the performance of a MOSFET is not affected by neutron bombardment while gamma irradiation affects its threshold voltage.

3.15 How do N-channel and P-channel MOSFETs differ in their response to gamma radiation. Which of these MOSFETs (NMOS or PMOS) is radiation hard, and why?

3.16 How does the output current of a MOSFET vary with temperature for operation near the threshold voltage? How does it change if the MOSFET is operating far away from it.

3.17 (a) In an N-channel IGBT with an N^+ polysilicon silicon gate, the gate oxide thickness (t_{ox}) is 800 Å, P-base doping concentration (N_A) is 1.5×10^{17} cm^{-3}, and gate doping is $N_D = 1.2 \times 10^{20}$ cm^{-3}. The interface state density is 1.7×10^{10} cm^{-2}. Find the threshold voltage of the IGBT.

(b) If N_A is raised to 5×10^{17} cm^{-3}, what is the new threshold voltage? (c) What is the effect of a higher oxide charge ($= 1 \times 10^{11}$ cm^{-2}) on the threshold voltage?

3.18 An N-channel power DMOSFET has linear cell geometry. It is working with a gate bias $V_{GS} = 8$ V at a drain–source voltage $V_{DS} = 5$ V. Its structural parameters are: polySi gate length (L_G) = 12 μm, cell pitch ($L_G + 2m$) = 25 μm, channel length (L_{Ch}) = 1.5 μm, gate oxide thickness (t_{ox}) = 900 Å, threshold voltage (V_{Th}) = 3.5 V, P-base depth (x_P) = 2.5 μm, resistivity of N$^-$-drift region (ρ_D) = 20 Ω-cm and its thickness (t) = 25 μm. Calculate the specific ON resistance of the MOSFET, taking the surface electron mobility value as 450 cm^2/V-sec.

3.19 The P-base concentration in an IGBT is 2.4×10^{17} cm^{-3} and the gate oxide thickness is 750 Å. Determine the threshold voltage at 253 K, 300 K, and 473 K. The values of intrinsic carrier concentration at these temperatures are 1×10^8 cm^{-3}, 1.5×10^{10} cm^{-3}, 2×10^{12} cm^{-3}, respectively, and the corresponding thermal voltages are 0.0218 V, 0.0259 V, 0.0322 V, respectively. You may neglect the work function difference and oxide charge effects.

REFERENCES

1. (a) J. E. Lilienfeld, Canadian Patent Application filed October 25; US Patent 1,745,175 (1930), 1,877,140 (1932), 1, 900, 018 (1933); (b) O. Heil, British Patent 439,457 (filed and granted 1935).
2. W. Shockley and G. L. Pearson, Modulation of Conductance of Thin Films of Semiconductors by Surface Charges, *The Physical Review*, The American Institute of Physics for the American Physical Society, Lancaster, Pennsylvania, Vol. 24, 1948, pp. 232–233.
3. H. Ikeda, K. Ashikawa, and K. Urita, Power MOSFETs for Medium-Wave and Short-Wave Transmitters, *IEEE Trans. Electron Devices*, Vol. 27, No. 2, February, 1980, pp. 330–334.
4. B. J. Baliga, Switching Lots of Watts at High Speed, *IEEE Spectrum*, Vol.18, 1981, pp. 42–48.
5. E. H. Nicollian and J. R. Brews, *MOS Physics and Technology*, Wiley-Interscience, New York, 1982.
6. Y. P. Tsividis, *Operation and Modelling of the MOS Transistor*, McGraw-Hill Book Co., New York, 1987.
7. D. A. Grant and J. Gowar, *Power MOSFETs: Theory and Applications*, John Wiley & Sons, New York, 1989.
8. M. N. Darwish and K. Board, Optimization of Breakdown voltage and ON-resistance of VDMOS Transistors, *IEEE Trans. Electron Devices*, Vol. ED-31, No. 12, December 1984, pp. 1769–1773.
9. S. C. Sun and J. D. Plummer, Modeling of the ON-resistance of LDMOS, VDMOS and VMOS Power Transistors, *IEEE Trans. Electron Devices*, Vol. ED-27, No. 2, February 1980, pp. 356–367.
10. S. D. Kim, I. J. Kim, M. Khan, and Y. I. Choi, An Accurate ON-resistance Model for Low-Voltage DMOS Devices, *Solid state Electronics*, Vol. 38, No. 2, February 1995, pp. 345–350.
11. C. Hu, M-. H. Chi, and V. M. Patel, Optimum Design of Power MOSFETs, *IEEE Trans. Electron Devices*, Vol. ED-31, No. 12, December 1984, pp. 1693–1700.

12. H. R. Chang, R. D. Black, V. A. K. Temple, W. Tantraporn, and B. J. Baliga, Self-Aligned UMOSFETs with Specific ON-resistance of 1 Milliohm-cm^2, *IEEE Trans. Electron Devices*, Vol. ED-34, No. 11, November 1987, pp. 2329–2334.
13. S. Matsumoto, T. Ohno, H. Ishii, and H. Yoshino, A High-Performance Self-Aligned UMOSFET with a Vertical Trench Contact structure, *IEEE Trans. Electron Devices*, Vol. 41, No.5, May 1994, pp. 814–818.
14. T. Syau, P. Venkatraman, and B. J. Baliga, Comparison of Ultralow Specific ON-resistance UMOSFET Structures: The ACCUFET, EXTFET, INVFET and Conventional UMOSFETs, *IEEE Trans. Electron Devices*, Vol. 41, No. 5, May 1994, pp. 800–808.
15. D. Ueda, H. Takagi and G. Kano, A New Vertical Power MOSFET Structure with Extremely Reduced On-Resistance, *IEEE Trans. Electron Devices*, Vol. ED-32, No. 1, January 1985, pp. 2–6.
16. B. J. Baliga in S. M. Sze (Ed.), *Modern Semiconductor Device Physics, Power Devices*, John Wiley & Sons, New York, 1998, pp. 183–252.
17. L. Vadasz and A. S. Grove, Temperature Dependence of MOS Transistor Characteristics Below Saturation, *IEEE Trans. Electron Devices*, Vol. ED-13, No. 12, December 1966, pp. 863–866.
18. R. Wang, J. Dunkley, T. A. DeMassa, and J. F. Jelsma, Threshold Voltage Variations with Temperature in MOS Transistors, *IEEE Trans. Electron Devices*, Vol. ED-18, No. 6, June 1971, pp. 386–388.
19. F. M. Klaassen and W. Hes, On the Temperature Coefficient of the MOSFET Threshold Voltage, *Solid State Electronics*, Vol. 29, No. 7, July 1986, pp. 787–789.

APPENDIX 3.1: DERIVATION OF EQS. (3.2a) AND (3.2b)

Electrons and holes in semiconductors obey the *Fermi–Dirac statistics*, which practically reduce to the mathematically simpler *Maxwell–Boltzmann statistics* in non-degenerate semiconductors:

$$f_{\mathrm{M}}(E) = \exp\{-(E - E_{\mathrm{F}})/kT\} \qquad \text{(A3.1.1)}$$

where E is the energy of the carrier (electron or hole), E_{F} is the Fermi energy level, k is the Boltzmann constant, and T is the absolute temperature.

In thermal equilibrium, the electron and hole concentrations are expressed as

$$n = N_{\mathrm{c}}\exp\{-(E_{\mathrm{C}} - E_{\mathrm{F}})/kT\},\ p = N_{\mathrm{V}}\exp\{-(E_{\mathrm{F}} - E_{\mathrm{V}})/kT\} \qquad \text{(A3.1.2)}$$

where N_{c} and N_{v} are the effective densities of states at the conduction and valence band edges.

In an intrinsic semiconductor, we have $n = p$. So, $E_{\mathrm{C}} - E_{\mathrm{F}} \approx E_{\mathrm{F}} - E_{\mathrm{v}}$. Then the Fermi level is approximately near the middle of the band gap. It is called the *intrinsic Fermi level* and is denoted by E_{i}. The intrinsic Fermi level is commonly used as a reference level for discussions on extrinsic semiconductors. As the intrinsic carrier concentration n_{i} is useful in relating the carrier concentrations in extrinsic semiconductors also, similarly E_i is used as a reference level in discussing extrinsic semiconductors. Because

$$n_i = N_c \exp\{-(E_C - E_i)/kT\} = N_V \exp\{-(E_i - E_V)/kT\} \quad \text{(A3.1.3)}$$

Eqs. (A3.1.2) are rewritten as

$$n = n_i \exp\{(E_F - E_i)/kT\}, \qquad p = n_i \exp\{-(E_F - E_i)/kT\} \quad \text{(A3.1.4)}$$

Focusing attention on the electrons, let $\Delta E_C = \Delta E_i = E_{ip} - E_{ip0}$ be the electron potential energy difference between a point on the surface of the semiconductor and another point inside the bulk semiconductor, at equilibrium. Then the electrostatic potential difference ψ between these two points is, by definition, the potential energy difference per unit charge given by

$$\psi = \frac{E_{ip} - E_{ip0}}{-q} \quad \text{(A3.1.5)}$$

Applying Eq. (A3.1.4), the electron concentrations at the two points considered are respectively

$$n_p = n_i \exp\{(E_F - E_{ip})/kT\}, \qquad n_{p0} = n_i \exp\{(E_F - E_{ip0})/kT\} \quad \text{(A3.1.6)}$$

which on division yield

$$\frac{n_p}{n_{p0}} = \exp\{-(E_{ip} - E_{ip0})/kT\} \quad \text{(A3.1.7)}$$

Since from Eq. (A3.1.5), we have $E_{ip} - E_{ip0} = -q\psi$, therefore Eq. (A3.1.7) reduces to

$$n_p = n_{p0} \exp\left(\frac{q\psi}{kT}\right) \quad \text{(A3.1.8)}$$

which is Eq. (3.2a). Similarly, starting from equation for hole density (p) in Eq. (A3.1.4), Eq. (3.2b) is obtained. This derivation is left as an exercise to the reader.

APPENDIX 3.2: DERIVATION OF EQ. (3.7)

Multiplying both sides of Eq. (3.6) with $d\psi/dx$, we get

$$\frac{d\psi}{dx} \cdot \frac{d^2\psi}{dx^2} = -\frac{q}{\varepsilon_0 \varepsilon_s}\left[p_{p0}\left\{\exp\left(-\frac{q\psi}{kT}\right) - 1\right\} - n_{p0}\left\{\exp\left(\frac{q\psi}{kT}\right) - 1\right\}\right] \cdot \frac{d\psi}{dx} \quad \text{(A3.2.1)}$$

Using the identity

$$\frac{1}{2}\frac{d}{dx}\left(\frac{d\psi}{dx}\right)^2 = \frac{d\psi}{dx}\cdot\frac{d^2\psi}{dx^2} \tag{A3.2.2}$$

we have

$$\begin{aligned}
&\frac{1}{2}\frac{d}{dx}\left(\frac{d\psi}{dx}\right)^2 \\
&= -\frac{q}{\varepsilon_0\varepsilon_s}\left[p_{p0}\left\{\exp\left(-\frac{q\psi}{kT}\right)-1\right\}-n_{p0}\left\{\exp\left(\frac{q\psi}{kT}\right)-1\right\}\right]\cdot\frac{d\psi}{dx} \\
&= -\frac{qp_{p0}}{\varepsilon_0\varepsilon_s}\left[\left\{\exp\left(-\frac{q\psi}{kT}\right)-1\right\}-\frac{n_{p0}}{p_{p0}}\left\{\exp\left(\frac{q\psi}{kT}\right)-1\right\}\right]\cdot\frac{d\psi}{dx} \\
&= -\frac{(kT/q)qp_{p0}}{\varepsilon_0\varepsilon_s(kT/q)}\left[\left\{\exp\left(-\frac{q\psi}{kT}\right)-1\right\}-\frac{n_{p0}}{p_{p0}}\left\{\exp\left(\frac{q\psi}{kT}\right)-1\right\}\right]\cdot\frac{d\psi}{dx} \\
&= \frac{(kT/q)}{\varepsilon_0\varepsilon_s(kT/q)/(qp_{p0})}\left[\left\{\exp\left(-\frac{q\psi}{kT}\right)-1\right\}-\frac{n_{p0}}{p_{p0}}\left\{\exp\left(\frac{q\psi}{kT}\right)-1\right\}\right]\cdot\frac{d\psi}{dx} \\
&= \frac{(kT/q)}{{L_D}^2}\left[\left\{\exp\left(-\frac{q\psi}{kT}\right)-1\right\}-\frac{n_{p0}}{p_{p0}}\left\{\exp\left(\frac{q\psi}{kT}\right)-1\right\}\right]\cdot\frac{d\psi}{dx} \qquad \text{(A3.2.3)}
\end{aligned}$$

where L_D is the extrinsic Debye length given by

$$L_D = \sqrt{\frac{\varepsilon_0\varepsilon_s kT}{q^2 p_{p0}}} \tag{A3.2.4}$$

Let

$$\frac{d\Omega}{dx} = \left[\left\{\exp\left(-\frac{q\psi}{kT}\right)-1\right\}-\frac{n_{p0}}{p_{p0}}\left\{\exp\left(\frac{q\psi}{kT}\right)-1\right\}\right]\cdot\frac{d\psi}{dx} \tag{A3.2.5}$$

Then

$$\Omega = \int\left[\left\{\exp\left(-\frac{q\psi}{kT}\right)-1\right\}-\frac{n_{p0}}{p_{p0}}\left\{\exp\left(\frac{q\psi}{kT}\right)-1\right\}\right]\cdot d\psi$$

$$= \frac{\exp\left(-\frac{q\psi}{kT}\right)}{-q/(kT)} - \psi - \frac{n_{p0}}{p_{p0}}\left\{\frac{\exp\left(\frac{q\psi}{kT}\right)}{q/(kT)} - \psi\right\} \quad \text{(A3.2.6)}$$

$$= \frac{-\exp\left(-\frac{q\psi}{kT}\right) - \left(\frac{q\psi}{kT}\right) - \frac{n_{p0}}{p_{p0}}\left\{\exp\left(\frac{q\psi}{kT}\right) - \left(\frac{q\psi}{kT}\right)\right\}}{q/(kT)}$$

$$= \frac{kT}{q}\left[-\exp\left(-\frac{q\psi}{kT}\right) - \left(\frac{q\psi}{kT}\right) - \frac{n_{p0}}{p_{p0}}\left\{\exp\left(\frac{q\psi}{kT}\right) - \left(\frac{q\psi}{kT}\right)\right\}\right]$$

From Eqs. (A3.2.3) and (A3.2.6), we can write

$$\frac{1}{2}\frac{d}{dx}\left(\frac{d\psi}{dx}\right)^2$$

$$= \frac{(kT/q)^2}{{L_D}^2}\cdot\frac{d}{dx}\left[-\exp\left(-\frac{q\psi}{kT}\right) - \left(\frac{q\psi}{kT}\right) - \frac{n_{p0}}{p_{p0}}\left\{\exp\left(\frac{q\psi}{kT}\right) - \left(\frac{q\psi}{kT}\right)\right\}\right] \quad \text{(A3.2.7)}$$

Integrating Eq. (A3.2.7) once, we obtain

$$\frac{1}{2}\left(\frac{d\psi}{dx}\right)^2\Bigg|_0^{x_d}$$

$$= \frac{(kT/q)^2}{{L_D}^2}\left[-\exp\left(-\frac{q\psi}{kT}\right) - \left(\frac{q\psi}{kT}\right) - \frac{n_{p0}}{p_{p0}}\left\{\exp\left(\frac{q\psi}{kT}\right) - \left(\frac{q\psi}{kT}\right)\right\}\right]\Bigg|_0^{x_d} \quad \text{(A3.2.8)}$$

The boundary conditions are stated as

$$\psi = \begin{cases} 0 \quad \text{and} \quad \dfrac{d\psi}{dx} = 0 & \text{at } x = x_d \\ \psi_s & \text{at } x = 0 \end{cases} \quad \text{(A3.2.9)}$$

where ψ_s stands for the surface potential and x_d for the space-charge layer width. Extracting the square root of both the sides, Eq. (A3.2.8) may be rewritten as

$$\frac{1}{\sqrt{2}}\frac{d\psi}{dx}\bigg|_0^{x_d} = \frac{(kT/q)}{L_D}\left[-\exp\left(-\frac{q\psi}{kT}\right) - \left(\frac{q\psi}{kT}\right) - \frac{n_{p0}}{p_{p0}}\left\{\exp\left(\frac{q\psi}{kT}\right) - \left(\frac{q\psi}{kT}\right)\right\}\right]^{0.5}\Bigg|_0^{x_d} \tag{A3.2.10}$$

or

$$\begin{aligned}
-\frac{d\psi}{dx}\bigg|_0^{x_d} &= \xi(x) \\
&= \frac{\sqrt{2}\,(kT/q)}{L_D}\left[\exp\left(-\frac{q\psi}{kT}\right) + \left(\frac{q\psi}{kT}\right)\right. \\
&\qquad \left. + \frac{n_{p0}}{p_{p0}}\left\{\exp\left(\frac{q\psi}{kT}\right) - \left(\frac{q\psi}{kT}\right)\right\}\right]^{0.5}\Bigg|_0^{x_d} \\
&= \frac{\sqrt{2}\,(kT/q)}{L_D}\left[\exp\left(-\frac{q\psi_s}{kT}\right) + \left(\frac{q\psi_s}{kT}\right) + \frac{n_{p0}}{p_{p0}}\left\{\exp\left(\frac{q\psi_s}{kT}\right) - \left(\frac{q\psi_s}{kT}\right)\right\}\right. \\
&\qquad \left. - \exp(-0) - (0) - \frac{n_{p0}}{p_{p0}}\{\exp(0) - (0)\}\right]^{0.5} \\
&= \frac{\sqrt{2}\,(kT/q)}{L_D}\left[\exp\left(-\frac{q\psi_s}{kT}\right) + \left(\frac{q\psi_s}{kT}\right)\right. \\
&\qquad \left. + \frac{n_{p0}}{p_{p0}}\left\{\exp\left(\frac{q\psi_s}{kT}\right) - \left(\frac{q\psi_s}{kT}\right)\right\} - 1 - \frac{n_{p0}}{p_{p0}}\right]^{0.5} \\
&= \frac{\sqrt{2}\,(kT/q)}{L_D}\left[\left\{\exp\left(-\frac{q\psi_s}{kT}\right) + \left(\frac{q\psi_s}{kT}\right) - 1\right\}\right. \\
&\qquad \left. + \frac{n_{p0}}{p_{p0}}\left\{\exp\left(\frac{q\psi_s}{kT}\right) - \left(\frac{q\psi_s}{kT}\right) - 1\right\}\right]^{0.5}
\end{aligned} \tag{A3.2.11}$$

APPENDIX 3.3: DERIVATION OF THE EQUATIONS FOR BULK SEMICONDUCTOR POTENTIAL ψ_B AND THE SURFACE CHARGE Q_S AT THE POINT OF TRANSITION INTO STRONG INVERSION

The relationship between the charge and electric field in a semiconductor is expressed by the Poisson's equation as

$$\frac{d\xi}{dx} = -\frac{d^2\psi}{dx^2} = -\frac{q}{\varepsilon_0 \varepsilon_s}\{(n-p)-(N_D-N_A)\} \qquad (A3.3.1)$$

where ξ is the electric field, ψ is the potential, q is the electronic charge, ε_0 is the permittivity of free space, ε_s is the dielectric constant of Si, n and p are the electron and hole concentrations, and N_D and N_A are the donor and acceptor densities.

The hole density p in P-type region is related to the Fermi level E_F in this region as

$$p = n_i \exp\left(\frac{E_i - E_F}{kT}\right) \qquad (A3.3.2)$$

where n_i is the intrinsic carrier concentration, E_i is intrinsic energy level, k is the Boltzmann constant, and T is the absolute temperature.

This equation follows from the formula

$$p = N_V \exp\left(-\frac{E_F - E_V}{kT}\right) \qquad (A3.3.3)$$

where N_v is the effective density of states in the valence band and E_v is the energy of the upper valence band edge. For an intrinsic semiconductor, $E_F = E_i$, so that

$$n_i = N_V \exp\left(-\frac{E_i - E_v}{kT}\right) \qquad (A3.3.4)$$

Substituting for N_v from Eq. (A3.3.4) into Eq. (A3.3.3) we get Eq. (A3.3.2).

The difference between the intrinsic and Fermi energy levels is defined as the *bulk semiconductor potential* $q\psi_B$. This enables us to rewrite Eq. (A3.3.2) as

$$p = n_i \exp\left(\frac{q\psi_B}{kT}\right) \qquad (A3.3.5)$$

Now, in the neutral P-region, the total space-charge density is zero. Hence

$$\frac{d\xi}{dx} = 0 \qquad (A3.3.6)$$

and

$$n - p - N_D + N_A = 0 \tag{A3.3.7}$$

In a P-type semiconductor, it may be assumed that $N_D = 0$ and $p \gg n$. Therefore, potential of the P-type semiconductor, designated as ψ_B, is obtained by setting $N_D = n = 0$ in Eq. (A3.3.7), yielding

$$p = N_A \tag{A3.3.8}$$

Combining Eqs. (A3.3.5) and (A3.3.8), we get

$$N_A = n_i \exp\left(\frac{q\psi_B}{kT}\right) \tag{A3.3.9}$$

from which

$$\psi_B = \frac{kT}{q}\ln\left(\frac{N_A}{n_i}\right) \tag{A3.3.10}$$

To derive the surface charge Q_s at the point of transition into strong inversion, we note that when $\psi_s < \psi_B$, the surface charge is from Eq. (3.9),

$$Q_s = \sqrt{2\varepsilon_0\varepsilon_s q p_{p0}\psi_s} \tag{A3.3.11}$$

The condition for strong inversion is $\psi_s > 2\psi_B$. Hence for the transition point, we write the following equation using Eq. (A3.3.10):

$$\psi_s = \frac{2kT}{q}\ln\left(\frac{N_A}{n_i}\right) \tag{A3.3.12}$$

The combination of Eqs. (A3.3.11) and (A3.3.12) gives

$$\begin{aligned} Q_s &= \sqrt{2\varepsilon_0\varepsilon_s p_{p0}\psi_s} = \sqrt{2\varepsilon_0\varepsilon_s q p_{p0}\frac{2kT}{q}\ln\left(\frac{N_A}{n_i}\right)} \\ &= \sqrt{4\varepsilon_0\varepsilon_s kTN_A\ln(N_A/n_i)} \end{aligned} \tag{A3.3.13}$$

APPENDIX 3.4: DERIVATION OF EQS. (3.33)-(3.36)

Equation (3.33) for Specific Channel Resistance of Linear Cell Geometry

Referring to Fig. 3.8, the length of the unit cell in the plane of the paper $= L_G + 2m$ while its length perpendicular to the plane of the paper =

channel width $= Z$. Hence, area of the unit cell is

$$A = (L_G + 2m)Z \tag{A3.4.1}$$

It is noted that in each unit cell, there are two channels, one in each half-side. Applying Eq. (3.23) and remembering that the area of cross section is doubled for the channels on the two halves of the unit cell, the total channel resistance is halved. Therefore,

$$r_{Ch} = \frac{L}{2Z\mu_{ns}C_{ox}(V_{GS} - V_{Th})} \tag{A3.4.2}$$

Multiplying by the area A from Eq. (A3.4.1), the specific channel resistance (i.e., resistance per unit area) is expressed as

$$r_{Ch,Sp} = \left\{\frac{L}{2Z\mu_{ns}C_{ox}(V_{GS} - V_{Th})}\right\} \times (L_G + 2m)Z = \frac{L(L_G + 2m)}{2\mu_{ns}C_{ox}(V_{GS} - V_{Th})} \tag{A3.4.3}$$

Equation (3.34) for Specific Accumulation Layer Resistance of Linear Cell Structure

The unit cell area A is given by Eq. (A3.4.1). Again referring to Fig. 3.8, the length of the accumulation region in the plane of the paper $= L_G - 2 \times$ (lateral diffusion of the P base). Taking the lateral diffusion of P base to be nearly equal to its vertical depth x_P, the above length $= L_G - 2x_P$. In the perpendicular direction, the length of the accumulation region is Z. From similar considerations to Eq. (A3.4.2), we can write the accumulation region resistance as

$$r_A = \frac{L_G - 2x_p}{2Z\mu_{nA}C_{ox}(V_{GS} - V_{Th})} \tag{A3.4.4}$$

where the inversion layer mobility μ_{ns} is replaced by the accumulation layer mobility μ_{nA}.

Now applying Eqs. (A3.4.1) and (A3.4.4), the specific accumulation layer resistance is

$$r_{A,Sp} = \frac{L_G - 2x_p}{2Z\mu_{nA}C_{ox}(V_{GS} - V_{Th})} \times (L_G + 2m)Z$$

$$= \frac{(L_G - 2x_p)(L_G + 2m)}{2\mu_{nA}C_{ox}(V_{GS} - V_{Th})} \tag{A3.4.5}$$

The approximation made in deriving this equation is the assumption that the current flows from the channel into the accumulation layer up to its center point and then flows downwards. In reality, the current spreads into the JFET region between the two P bases from the channels in the two halves of the unit cell. So, the two-dimensional nature of current spreading from the channels through the accumulation layer into the JFET region must be taken into consideration. This is reasonably accounted for by multiplying the specific accumulation layer resistance by a factor K. A value of $K = 0.6$ gives good agreement between theoretical calculations and experimental results. Based on these arguments, Eq. (A3.4.5) is written in the revised form as

$$r_{\mathrm{Ch,A}} = \frac{0.6(L_{\mathrm{G}} - 2x_{\mathrm{p}})(L_{\mathrm{G}} + 2m)}{2\mu_{\mathrm{nA}}C_{\mathrm{ox}}(V_{\mathrm{GS}} - V_{\mathrm{Th}})} \qquad \text{(A3.4.6)}$$

Equation (3.35) for Specific JFET Region Resistance of Linear Cell Structure

As before, length of the JFET region horizontally in the plane of the paper $= L_{\mathrm{G}} - 2x_{\mathrm{P}} - 2 \times$ Depletion layer thickness extending under the gate $= L_{\mathrm{G}} - 2x_{\mathrm{P}} - 2W_0$. The current flows vertically downwards a distance = Junction depth of P base + Depletion layer thickness $= x_{\mathrm{P}} + W_0$, in the plane of the paper. In the orthogonal direction, the JFET region stretches up to a distance $= Z$. Therefore, resistance of the JFET region is

$$r_{\mathrm{JFET}} = \frac{\text{Resistivity of the drift region} \times \text{length}}{\text{Cross-sectional area}} = \frac{\rho_{\mathrm{D}} \times (x_{\mathrm{p}} + W_0)}{(L_{\mathrm{G}} - 2x_{\mathrm{p}} - 2W_0)Z} \qquad \text{(A3.4.7)}$$

Multiplying by the unit cell area from Eq. (A3.4.1), the specific JFET region resistance is

$$r_{\mathrm{JFET,Sp}} = \frac{\rho_{\mathrm{D}} \times (x_{\mathrm{p}} + W_0)}{(L_{\mathrm{G}} - 2x_{\mathrm{p}} - 2W_0)Z} \times (L_{\mathrm{G}} + 2m)Z$$

$$= \frac{\rho_{\mathrm{D}}(L_{\mathrm{G}} + 2m)(x_{\mathrm{p}} + W_0)}{(L_{\mathrm{G}} - 2x_{\mathrm{p}} - 2W_0)} \qquad \text{(A3.4.8)}$$

Equation (3.36) for Specific Drift Region Resistance of Linear Cell Structure

The drift region resistance for half unit cell is given by the expression

$$R_{\mathrm{D}} = \frac{\rho_{\mathrm{D}}}{Z} \ln\left(\frac{a + t}{a}\right) \qquad \text{(A3.4.9)}$$

Considering both halves of the unit cell, we have

$$R_{\mathrm{D}} = \frac{\rho_{\mathrm{D}}}{2Z} \ln\left(\frac{a+t}{a}\right) \tag{A.3.4.10}$$

Multiplying the equation for R_{D} by the unit cell area, Eq. (3.4.1), the drift region resistance per square centimeter is

$$r_{\mathrm{D,Sp}} = \frac{\rho_{\mathrm{D}}}{2Z} \ln\left(\frac{a+t}{t}\right) \times (L_{\mathrm{G}} + 2m)Z = \frac{\rho_{\mathrm{D}}}{2}(L_{\mathrm{G}} + 2m) \ln\left(\frac{a+t}{t}\right) \tag{A3.4.11}$$

APPENDIX 3.5: DERIVATION OF EQ. (3.39)

Looking at Fig. 3.8, the resistance to current flow is modelled as the sum of two components: the first component due to current spreading at 45° angle from a width W_{t} to the total unit cell width $= W_{\mathrm{m}} + W_{\mathrm{t}}$, and the second component due to current flowing in a uniform unit cell width $= W_{\mathrm{m}} + W_{\mathrm{t}}$. For the first component, the cross-sectional area is $= W_{\mathrm{t}} \times Z$, where Z is the width of the unit cell in the direction perpendicular to the plane of the drawing. For the second component, the cross-sectional area is $= (W_{\mathrm{m}} + W_{\mathrm{t}}) \times Z$, where Z is the width of the unit cell in the direction perpendicular to the plane of the drawing.

To evaluate the first component of this resistance, the resistance offered to the current in a region of increasing cross-sectional area is calculated. The schematic drawing of the model for calculation of ON-resistance is shown in Fig. A3.5.1. Uniform carrier flow is assumed from the accumulation layer formed under the gate-drain overlap to the N^+ drain region enclosed by the hatched polygonal area in the diagram. W_0 is the zero-bias depletion region width. For one-half unit cell (marked by AOCDH), the resistance due to the first component is

$$\rho_{\mathrm{D}} \int_0^{W_{\mathrm{m}}/2} \frac{dx}{\left(\frac{W_{\mathrm{t}}}{2} + x\right)Z} = \frac{\rho_{\mathrm{D}}}{Z}\left[\ln\left(\frac{W_{\mathrm{t}}}{2} + x\right)\right]_0^{W_{\mathrm{m}}/2}$$

$$= \frac{\rho_{\mathrm{D}}}{Z}\left\{\ln\left(\frac{W_{\mathrm{m}} + W_{\mathrm{t}}}{2}\right) - \ln\left(\frac{W_{\mathrm{t}}}{2}\right)\right\} = \frac{\rho_{\mathrm{D}}}{Z} \ln\left(\frac{W_{\mathrm{m}} + W_{\mathrm{t}}}{W_{\mathrm{t}}}\right) \tag{A3.5.1}$$

The upper limit of integration is $W_{\mathrm{m}}/2$ because the spreading angle is 45°. When the current spreads up to a distance $\mathrm{AO} = W_{\mathrm{m}}/2$ horizontally, the vertical depth OB is also $W_{\mathrm{m}}/2$ since $\tan 45° = \mathrm{OB}/\mathrm{AO} = 1$ or, $\mathrm{OB} = \mathrm{AO} =$

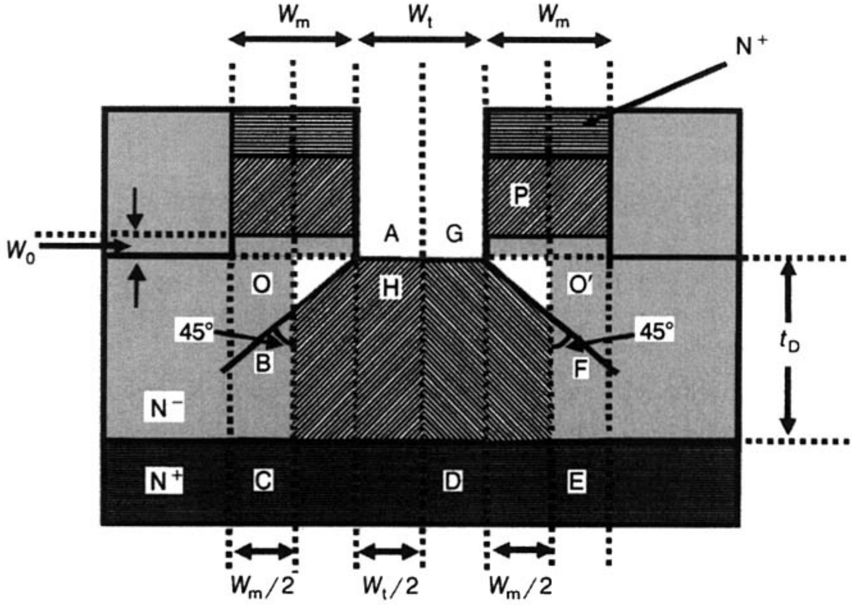


Figure A3.5.1 Model for calculation of spreading resistance.

$W_m/2$. For the other half unit cell (marked by GO′EDH), the same expression applies so that the resistance of the full unit cell is half of the above:

$$\frac{\rho_D}{2Z}\ln\left(\frac{W_m+W_t}{W_t}\right) \qquad \text{(A3.5.2)}$$

Since the area of unit cell = $(W_m + W_t)Z$, the specific resistance of unit cell because of the first component is

$$R_1 = \frac{\rho_D}{2Z}\ln\left(\frac{W_m+W_t}{W_t}\right) \times (W_m+W_t)Z$$

$$= \frac{\rho_D(W_m+W_t)}{2}\ln\left(1+\frac{W_m}{W_t}\right) \qquad \text{(A3.5.3)}$$

After the current has traversed a depth = $W_m/2$ vertically downwards, the remaining thickness of N$^-$ base is $t_D - W_m/2$. Hence the resistance due to second component is

$$\frac{\rho_D\left(t_D - \frac{W_m}{2}\right)}{(W_m + W_t)Z} \tag{A3.5.4}$$

from which the specific resistance for the second component is

$$R_2 = \frac{\rho_D\left(t_D - \frac{W_m}{2}\right)}{(W_m + W_t)Z} \times (W_m + W_t)Z = \rho_D\left(t_D - \frac{W_m}{2}\right) \tag{A3.5.5}$$

The total specific resistance is the series combination of R_1 and R_2 given by

$$\begin{aligned} R = R_1 + R_2 &= \frac{\rho_D(W_m + W_t)}{2}\ln\left(1 + \frac{W_m}{W_t}\right) + \rho_D\left(t_D - \frac{W_m}{2}\right) \\ &= \rho_D\left\{\frac{(W_m + W_t)}{2}\ln\left(1 + \frac{W_m}{W_t}\right) + \left(t_D - \frac{W_m}{2}\right)\right\} \end{aligned} \tag{A3.5.6}$$

This equation holds as long as t_D is $\geq W_m/2$. When t_D is $\leq W_m/2$, the term $\rho_D(t_D - W_m/2)$ is removed because the current reaches the drain contact before spreading fully up to the edge of the cell at the 45° angle. Then $R = R_1$.

APPENDIX 3.6: DERIVATION OF EQ. (3.49)

The bulk semiconductor potential is given by Eq. (A3.3.10)

$$\psi_B = \left(\frac{kT}{q}\right)\ln\left(\frac{N_A}{n_t}\right) \tag{A3.6.1}$$

where N_A is the doping concentration of P-type Si and n_i is the intrinsic carrier concentration.

Square of the intrinsic carrier density is expressed as

$$n_i^2 = 1.5 \times 10^{33} \times T^3 \times \exp\left(-\frac{E_{G0}}{kT}\right) \tag{A3.6.2}$$

where T is the absolute temperature and E_{G0} is the energy gap of Si at 0 K (1.21 eV). So,

$$n_i = 3.87 \times 10^{16} \times T^{1.5} \times \exp\left(-\frac{E_{G0}}{2kT}\right) \tag{A3.6.3}$$

Substituting for n_i from Eq. (A3.6.3) in Eq. (A3.6.1), we get

$$\psi_{\mathrm{B}} = \left(\frac{kT}{q}\right) \ln \left\{ \frac{N_{\mathrm{A}}}{3.87 \times 10^{16} \times T^{1.5} \times \exp\left(-\frac{E_{\mathrm{G0}}}{2kT}\right)} \right\}$$

$$= \left(\frac{kT}{q}\right) \ln \left\{ C \times T^{-1.5} \times \exp\left(\frac{E_{\mathrm{G0}}}{2kT}\right) \right\} \qquad \text{(A3.6.4)}$$

where C is a constant $= N_{\mathrm{A}}/(3.87 \times 10^{16})$. Differentiating both sides of Eq. (A3.6.4) with respect to T, we have

$$\frac{d\psi_{\mathrm{B}}}{dT} = \left(\frac{k}{q}\right) \ln \left\{ C \times T^{-1.5} \times \exp\left(\frac{E_{\mathrm{G0}}}{2kT}\right) \right\} + \left(\frac{kT}{q}\right) \left[\frac{1}{\left\{ C \times T^{-1.5} \times \exp\left(\frac{E_{\mathrm{G0}}}{2kT}\right) \right\}} \right.$$

$$\times C \left\{ -1.5 \times T^{-2.5} \times \exp\left(\frac{E_{\mathrm{G0}}}{2kT}\right) + T^{-1.5} \right.$$

$$\left. \left. \times \exp\left(\frac{E_{\mathrm{G0}}}{2kT}\right) \times \left(\frac{E_{\mathrm{G0}}}{2k}\right) \times -1 \times T^{-2} \right\} \right]$$

$$= \frac{1}{T}\psi_{\mathrm{B}} + \left(\frac{kT}{q}\right) \times \left\{ -1.5 \times T^{-1} - T^{-2} \times \left(\frac{E_{\mathrm{G0}}}{2k}\right) \right\} \qquad \text{(A3.6.5)}$$

using Eq. (A3.6.4) for ψ_{B}. Equation (A3.6.5) may be rewritten as

$$\frac{d\psi_{\mathrm{B}}}{dT} = \frac{1}{T}\psi_{\mathrm{B}} - \frac{1}{T}\left(\frac{E_{\mathrm{G0}}}{2q}\right) - \frac{1}{T} \times 1.5\left(\frac{kT}{q}\right) \qquad \text{(A3.6.6)}$$

Neglecting the term involving 1.5 (kT/q), this equation may be approximately expressed in the form

$$\frac{d\psi_{\mathrm{B}}}{dT} \cong \pm \frac{1}{T}\left\{ \left(\frac{E_{\mathrm{G0}}}{2q}\right) - |\psi_{\mathrm{B}}| \right\} \qquad \text{(A3.6.7)}$$

4

BIPOLAR COMPONENTS OF IGBT

The IGBT cross section contains three P-N junctions (N^+-P, P-N^-, and N^--P^+), one P-I-N diode (as P^+-N^-) two transistor sections (N^+-P-N^- and P-N^--P^+), and one thyristor (N^--P-N^--P^+). Sandwiched between any two neighboring IGBT cells is the junction field effect transistor (JFET) region. In this chapter we shall develop concepts of bipolar devices, which will be useful to us in our study of IGBT.

4.1 PN JUNCTION DIODE

Power diodes [1–4] are manufactured in current ratings from a few amperes to several hundred amperes and in voltage ratings from tens of volts to several thousand volts. Current ratings are valid only if diodes are mounted on suitable heat sinks so that temperature does not exceed a specified limit. Besides average current, RMS current, and repetitive peak current, a surge current is given, indicating the diode's capability to withstand an occasional transient or circuit fault. Both repetitive and nonrepetitive reverse voltages are specified. An important parameter is the reverse recovery time, defined as the interval after which the diode recovers its ability to block reverse voltage, measured from the instant the forward conduction has stopped. On this criterion, fast recovery and slow recovery diodes, are distinguished.

The Insulated Gate Bipolar Transistor (IGBT) Theory and Design, By Vinod Kumar Khanna.
ISBN 0-471-23845-7 © 2003 Institute of Electrical and Electronics Engineers

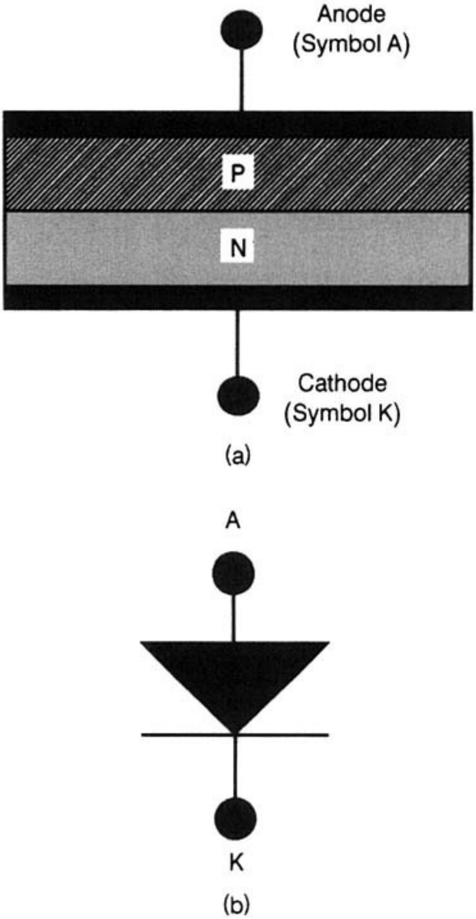


Figure 4.1 P-N diode: (a) Junction structure and (b) circuit diagram symbol.

The diode has two terminals, Fig. 4.1a, with the anode A making contact with the P-side surface and cathode K contacting the N-side surface. Forward current flows from anode (A) to cathode (K). Figure 4.1b shows the circuit symbol of the device. The current rating is determined by the diode area; the voltage rating is decided by the resistivity, doping profile, and thickness of starting N-type wafer. Additionally, the voltage withstanding capability is determined by chip edge termination and passivation for surface electric field control. The reverse recovery time and forward drop are controlled both by the impurity profile and carrier lifetime in N-region.

Diodes are mounted in a variety of packages, the same as those for small-signal transistors up to high power transistors and thyristors. There are two major types of packages for power diodes: stud-type package and hockey-puck package (Fig. 4.2). In the *stud-type package*, the metal casing and

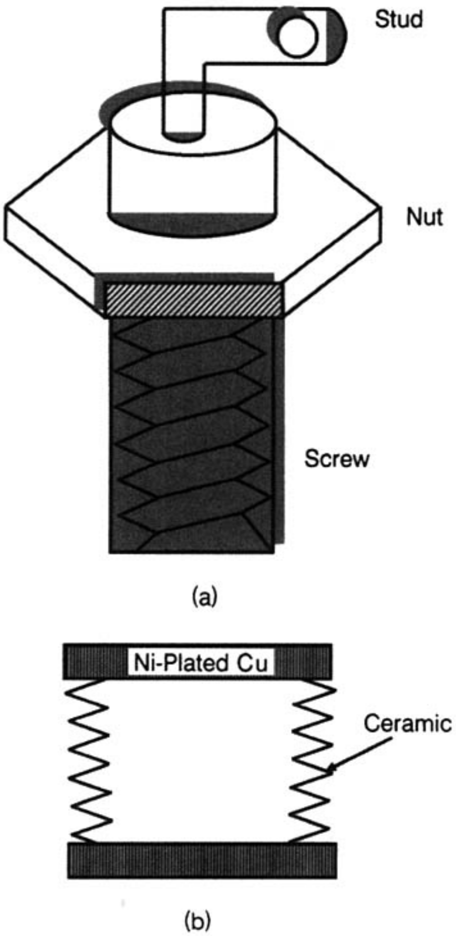


Figure 4.2 Commonly used packages of a power diode: (a) Stud-type package and (b) disk-type or hockey-puck package.

the stud constitute one terminal of the diode while the other terminal is the on opposite side, properly insulated from the casing. The stud can have either polarity—A or K. The casing can be easily fixed on the heat sink. In the *hockey-puck package*, the two terminals of the device are flat metallic surfaces isolated by ceramic insulator. It is specially suited for high current and voltage ratings.

4.1.1 Built-in Potential (ϕ_0)

Figure 4.3 represents the physical model of an abrupt PN junction in which the impurity concentration changes from N_D donors in the N-type semiconductor to N_A acceptors in P-type semiconductor [4, 5]. Electron diffusion

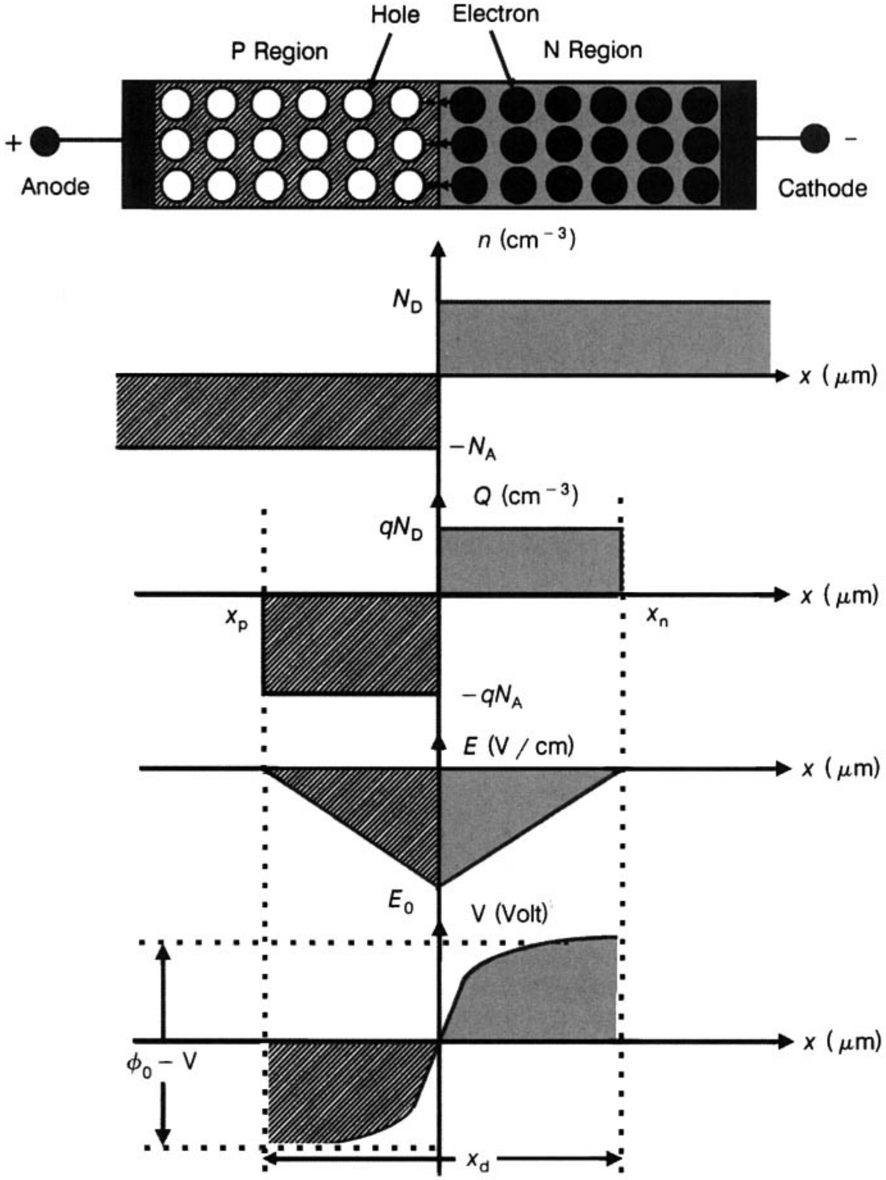

Hole
Electron
P Region
N Region
+
Anode
-
Cathode
n (cm⁻³)
N_D
x (μm)
−N_A
Q (cm⁻³)
qN_D
x (μm)
x_p
x_n
−qN_A
E (V / cm)
x (μm)
E_0
V (Volt)
x (μm)
φ_0 − V
x_d

Figure 4.3 Physical model of an abrupt PN junction in which the impurity concentration changes from N_D donors in the N-type semiconductor to N_A acceptors in P-type semiconductor.

from N-type material into P-type material leaves behind positively charged fixed ions in the N-type material near the junction. Similarly, hole diffusion from P-type material into N-type material leaves behind negatively charged fixed ions in the P-type material, also near the junction. Uncovering of positive and negative charges by diffusion of free carriers across the junction develops an electric field, which creates carrier motion in the opposite direction. The PN junction attains equilibrium when the drift and diffusion currents cancel each other. Let x_p represent the distance over which the donor atoms acquire a positive charge due to loss of a free electron and let x_n be the distance over which the acceptor atoms have a negative charge due to loss of a hole. For definitions of the various symbols, please see Table 4.1. Then the region around the metallurgical junction, which is depleted of free carriers, called the *depletion* or *space-charge region*, is

$$x_\mathrm{d} = x_\mathrm{n} - x_\mathrm{p} \tag{4.1}$$

For a PN junction in equilibrium, we can express the hole current density [Appendix 4.1, Eq. (A4.1.13)] as

$$J_\mathrm{p}(x) = q\left\{ \mu_\mathrm{p} p(x) \xi(x) - D_\mathrm{p}\frac{dp(x)}{dx} \right\} \tag{4.2}$$

where q is the electronic charge, μ_p is the hole mobility, $p(x)$ is the hole concentration, $\xi(x)$ is the electric field, and D_p is the diffusion coefficient of holes. The first term of this equation represents the hole drift current density, and the second term represents the hole diffusion current density. Under equilibrium, $J_\mathrm{p}(x) = 0$; so, Eq. (4.2) reduces to

$$\left(\frac{\mu_\mathrm{p}}{D_\mathrm{p}}\right)\xi(x) = -\left(\frac{\mu_\mathrm{p}}{D_\mathrm{p}}\right)\frac{dV(x)}{dx} = \left\{\frac{1}{p(x)}\right\}\frac{dp(x)}{dx} \tag{4.3}$$

using Einstein's relation (Appendix 4.2) connecting the mobility with diffusion coefficient: $D_\mathrm{p}/\mu_\mathrm{p} = kT/q$ where k is Boltzmanns constant and T is the temperature in the Kelvin scale. Also, the electric field has been expressed as the negative differential coefficient of the electrostatic potential V as $\xi(x) = -dV(x)/dx$.

Integrating for potential V from P to N side having potentials ϕ_p and ϕ_n, respectively, we get

$$-\frac{q}{kT}\int_{\phi_\mathrm{p}}^{\phi_\mathrm{n}} dV = \int_{p_\mathrm{p}}^{p_\mathrm{n}} \frac{dp}{p} \tag{4.4}$$

or

$$-\frac{q}{kT}(\phi_\mathrm{n} - \phi_\mathrm{p}) = \ln\left(\frac{p_\mathrm{n}}{p_\mathrm{p}}\right) \tag{4.5}$$

Table 4.1 Glossary of Terms

Symbol	Definition	Unit
A	Junction area, chip area	cm^2
A	Impurity gradient	cm^{-4}
C_j	Depletion layer or junction capacitance of diode	pF, F
D_a	Ambipolar diffusion constant	cm^2/sec
D_B, D_C, D_E	Diffusion constants of minority carriers in the base, collector and emitter	cm^2/sec
D_n	Diffusion constant of electrons	cm^2/sec
D_p	Diffusion constant of holes	cm^2/sec
D	Half-width of intrinsic region of a PIN diode	cm, μm
E_g	Energy gap at room temperature	eV
E_0	Electric field at the junction	V/cm
E_{max}, E_m	Maximum electric field	V/cm
ξ	Electric field	V/cm
F	Function	No unit
h_{FE}	Common-emitter current gain of transistor	No unit
I_A	Anode current of thyristor	A
I_B, I_C, I_E	Base, collector, and emitter currents of a transistor	A
I_{CO}	Collector current with emitter open	A
i_D	Diode current	A
I_{DS}	Drain–source current	A
I_E	Emitter current of a transistor	A
I_G	Gate current of thyristor	A
I_{gen}	Generation current	A
I_K	Cathode current of thyristor	A
I_{nE}	Electron component of emitter current	A
I_{pE}	Hole component of emitter current	A
I_R	Reverse leakage current of a diode	A
I_s	Saturation current of the diode	A
J	Current density	A/cm^2
J_n, J_p	Electron, hole current density	A/cm^2
$J_n(0), J_p(0)$	Electron, hole current density at $x = 0$	A/cm^2
k	Boltzmann's constant	J/K
L_a	Ambipolar diffusion length	cm, μm
L_B, L_C, L_E	Diffusion lengths of minority carriers in the base, collector, and emitter	cm^2/sec
L_n, L_p	Electron, hole diffusion length	cm, μm
M	Collector multiplication factor	No unit
n, p	Electron, hole concentration	cm^{-3}
N_A, N_D	Acceptor, donor doping concentration	cm^{-3}
N_B	Background doping concentration	cm^{-3}
n_C, n_E	Equilibrium densities of minority carriers in the collector and emitter	cm^{-3}
n_i	Intrinsic carrier concentration	cm^{-3}
n_{p0}	Electron concentration on the P side	cm^{-3}
$n_p(0)$	Electron concentration on the P side at $x = 0$	cm^{-3}

Table 4.1 (*Continued*)

Symbol	Definition	Unit
Q_j	Depletion charge	C
Q	Electronic charge	C
P	Hole concentration	cm^{-3}
p_B	Equilibrium hole concentration in the base	cm^{-3}
p_n	Hole concentration on the N side	cm^{-3}
p'_n	Excess hole concentration on the N side over thermal equilibrium value	cm^{-3}
p_{n0}	Equilibrium hole concentration on the N side	cm^{-3}
$p_n(0)$	Hole concentration on the N side at $x = 0$	cm^{-3}
p_p	Hole concentration on the P side	cm^{-3}
T	Absolute temperature	K
T	Total N^--base thickness	cm, μm
t_d, t_r	Delay and rise times	μsec
V	Potential difference	V
v_a	External applied voltage	V
V_B	Breakdown voltage	V
V_{CB}	Collector–base voltage	V
V_{EB}	Emitter–base voltage	V
V_{N-}	Voltage drop in the intrinsic region of PIN diode	V
V_{N+}	Voltage drop in the end N^+ region of PIN diode	V
V_{P+}	Voltage drop in the end P^+ region of PIN diode	V
W	Depletion layer width	cm, μm
x_d	Total depletion layer width	cm, μm
x_n	Depletion layer width on the N side	cm, μm
x_p	Depletion layer width on the P side	cm, μm
α	Common-base current gain	No unit
α_1	Common-base current gain of the PNP transistor of thyristor	No unit
α_2	Common-base current gain of the NPN transistor of thyristor	No unit
α_T	Base transport factor	No unit
β	Common-emitter current gain	No unit
γ	Emitter injection efficiency	No unit
ε_0	Permittivity of free space	F/cm
ε_s	Relative permittivity of silicon	No unit
μ_n, μ_p	Electron, hole mobilities	cm^2/V-sec
ϕ	Potential, thermal voltage (kT/q)	V
ϕ_0	Built-in potential	V
ϕ_n, ϕ_p	Potentials on the N and P sides	V
τ_{HL}	High-level lifetime	μsec
τ_p	Hole lifetime	μsec
τ_{sc}	Space-charge generation lifetime	μsec

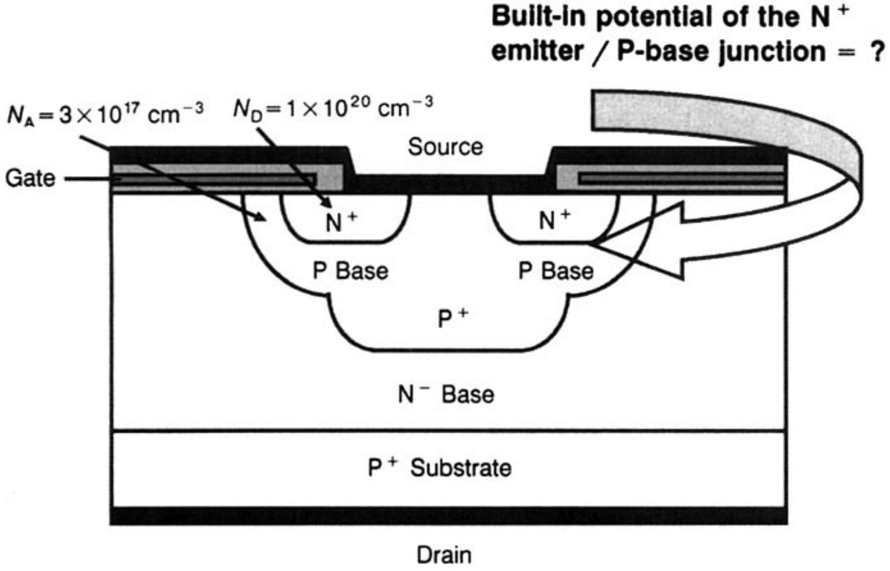

Built-in potential of the N$^+$ emitter / P-base junction = ?
$N_A = 3 \times 10^{17}$ cm^{-3}
$N_D = 1 \times 10^{20}$ cm^{-3}
Source
Gate
N$^+$
N$^+$
P Base
P Base
P$^+$
N$^-$ Base
P$^+$ Substrate
Drain

Figure E4.1.1 IGBT structure for Example 4.1.

because $\int dx/x = \ln x$. The symbols p_p and p_n denote the hole concentrations on the P and N sides, respectively. $p_p = N_A$ and because in a semiconductor under thermal equilibrium the product of electron and hole concentrations equals the square of the intrinsic carrier density, $p_n = n_i^2/N_D$ where n_i is the intrinsic carrier concentration, this gives

$$\phi_n - \phi_p = \phi_0 = \frac{kT}{q}\ln\left(\frac{p_p}{p_n}\right) = \frac{kT}{q}\ln\left(\frac{N_A N_D}{n_i^2}\right) \tag{4.6}$$

where ϕ_0 is the barrier or built-in potential.

Example 4.1 The N$^+$ emitter concentration of an IGBT is 1×10^{20} cm^{-3} and the P-base concentration is 3×10^{17} cm^{-3}. What is the built-in potential of the N$^+$ emitter/P-base junction?

Figure E4.1.1 illustrates the IGBT. Built-in potential $\phi_0 = (kT/q)\ln(N_A N_D/n_i^2) = 0.0259 \ln\{(3 \times 10^{17} \times 1 \times 10^{20})/(1.45 \times 10^{10})^2\} = 1.023$ V.

4.1.2 Depletion-Layer Width (x_d) and Capacitance (C_j)

By charge neutrality condition, the charges on both sides of the junction must be equal, that is,

$$qN_D x_n = -qN_A x_p \tag{4.7}$$

where q is the electronic charge. The electric field in the depletion region is determined by integrating the negative depletion charge concentration. Applying the integral form of Gauss' law, the peak electric field (E_0), which occurs at the junction, is given by

$$E_0 = \frac{qN_A x_p}{\varepsilon_0 \varepsilon_s} = -\frac{qN_D x_n}{\varepsilon_0 \varepsilon_s} \tag{4.8}$$

where ε_s is the relative permittivity of Si. When an external voltage v_a is applied to a PN junction, the potential drop across the depletion region is found by integrating the negative electric field, leading to

$$\phi_0 - v_a = -E_0\left(\frac{x_n - x_p}{2}\right) \tag{4.9}$$

where

$$\phi_0 = \frac{kT}{q}\ln\left(\frac{N_A N_D}{n_i^2}\right) \tag{4.10}$$

The symbol k denotes Boltzmann's constant, and n_i represents the intrinsic carrier concentration. Simultaneous solution of Eqs. (4.7), (4.8), and (4.9) gives the widths of the depletion region in N-type and P-type regions as

$$x_n = \sqrt{\frac{2\varepsilon_0 \varepsilon_s(\phi_0 - v_a)N_A}{qN_D(N_A + N_D)}} \tag{4.11}$$

and

$$x_p = -\sqrt{\frac{2\varepsilon_0 \varepsilon_s(\phi_0 - v_a)N_D}{qN_A(N_A + N_D)}} \tag{4.12}$$

from which the total depletion region width is

$$x_d = x_n + x_p = \sqrt{\frac{2\varepsilon_0 \varepsilon_s(\phi_0 - v_a)(N_A + N_D)}{qN_A N_D}} \tag{4.13}$$

x_d varies as the square root of the difference between built-in potential and externally applied voltage. Also, $x_d = x_n$ when $N_A \gg N_D$ and $x_d = x_p$ when $N_D \gg N_A$.

The dipole formed by the uncovered fixed ionic charges adjoining the junction constitutes a capacitance C_j, called the *depletion-layer capacitance.* It is given by

$$C_j = \frac{dQ_j}{dv_a} \tag{4.14}$$

where Q_j is the depletion charge expressed as

$$Q_j = |qAN_A x_p| = qAN_D x_n = A\sqrt{\frac{2q\varepsilon_0\varepsilon_s N_A N_D(\phi_0 - v_a)}{N_A + N_D}} \tag{4.15}$$

from which C_j is found to be

$$C_j = A\sqrt{\frac{q\varepsilon_0\varepsilon_s N_A N_D}{2(N_A + N_D)(\phi_0 - v_a)}} \tag{4.16}$$

4.1.3 Breakdown Voltage (V_B)

From Eqs. (4.8) and (4.11) or (4.12), the magnitude of electric field at the junction is

$$E_0 = \sqrt{\frac{2qN_A N_D(\phi_0 - v_a)}{\varepsilon_0\varepsilon_s(N_A + N_D)}} \tag{4.17}$$

The breakdown voltage (V_B) of a reverse-biased junction is governed by the maximum electric field $E_{max} = 3 \times 10^5$ V/cm which can be supported by the depletion region,

$$V_B \approx \frac{\varepsilon_0\varepsilon_s(N_A + N_D)E_{max}{}^2}{2qN_A N_D} \tag{4.18}$$

where it has been assumed that $|v_a| \gg \phi_0$. If $N_A \gg N_D = N_B$,

$$V_B \approx \frac{\varepsilon_0\varepsilon_s E_{max}{}^2}{2qN_B} \tag{4.19}$$

where N_B is the ionized background doping density of the lightly doped side in cm^{-3}. Substituting the values, of the parameters in Eq. (4.19), we have

$$V_B = \frac{2.96 \times 10^{17}}{N_D} \tag{4.20}$$

For Si, the maximum electric field E_m is written as

$$E_m = \frac{4 \times 10^5}{1 - \frac{1}{3}\log_{10}\left(\frac{N_B}{10^{16}}\right)} \quad \text{(volts/cm)} \tag{4.21}$$

Since the avalanche process involves band-to-band excitations, the breakdown voltage increases with energy bandgap. An approximate universal equation for the avalanche breakdown voltage of abrupt junction diodes (valid for Si, Ge, and GaAs) is

$$V_B = 60\left(\frac{E_g}{1.1}\right)^{3/2}\left(\frac{N_B}{10^{16}}\right)^{-3/4} \quad \text{(volts)} \qquad (4.22)$$

where E_g is the room-temperature energy gap. For linearly graded junctions,

$$V_B = 60\left(\frac{E_g}{1.1}\right)^{6/5}\left(\frac{a}{3\times10^{20}}\right)^{-2/5} \quad \text{(volts)} \qquad (4.23)$$

where a is the impurity gradient in cm^{-4}.

Example 4.2 In a nonpunchthrough IGBT, the impurity dopant density of the N$^-$ base is 1×10^{14} cm^{-3} while that of the P base is 2×10^{17} cm^{-3}. Calculate the (a) breakdown voltage using both Eqs. (4.20) and (4.22); (b) the depletion layer width on N side, (c) the depletion layer width on P side, (d) the total depletion layer width and (e) the undepleted length of the N$^-$ base at 1200 V given that the total N$^-$-base thickness is 200 μm.

The nonpunchthrough IGBT is shown in Fig. E4.2.1

(a) $V_B = 2.96\times10^{17}/N_D = 2.96\times10^{17}/(1\times10^{14}) = 2960$ V. Also, $V_B = 60(E_g/1.1)^{3/2}(N_B/10^{16})^{-3/4} = 60(1\times10^{14}/10^{16})^{-3/4} = 1897.4$ V.

(b) $x_n = [\{2\varepsilon_0\varepsilon_s(\phi_0 - v_a)N_A\}/\{qN_D(N_A+N_D)\}]^{1/2} = [\{2\times8.854\times10^{-14}\times11.9\times1200\times2\times10^{17}\}/\{1.6\times10^{-19}\times1\times10^{14}\times(2\times10^{17}+1\times10^{14})\}]^{1/2} = 0.012568$ cm $= 125.7$ μm.

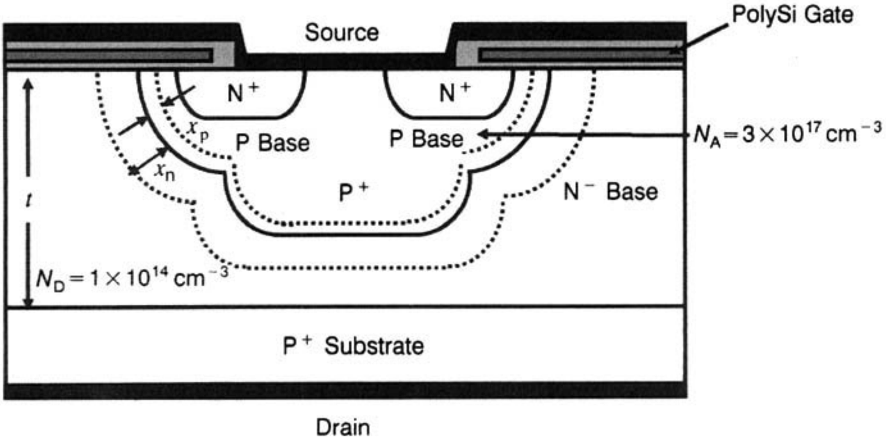


Figure E4.2.1 IGBT structure for Example 4.2.

(c) $x_p = [\{2\varepsilon_0\varepsilon_s(\phi_0 - v_a)N_D\}/\{qN_A(N_A + N_D)\}]^{1/2} = [\{2 \times 8.854 \times 10^{-14} \times 11.9 \times 1200 \times 1 \times 10^{14}\}/\{1.6 \times 10^{-19} \times 2 \times 10^{17} \times (2 \times 10^{17} + 1 \times 10^{14})\}]^{1/2} = 6.28 \times 10^{-6}$ cm $= 6.3 \times 10^{-2}$ μm.

(d) $x_d = x_n + x_p = 125.7$ μm $+ 6.3 \times 10^{-2} = 125.76$ μm.

(e) If t is the total N$^-$-base thickness, undepleted N$^-$-base width is $x = t - x_d = 200 - 125.76 = 74.24$ μm.

Example 4.3 Given that the breakdown voltage V_B of a P$^+$N diode is inversely proportional to the 3/4ths power of the doping concentration N_D of the lightly doped side, and minimum width of drift region (W_d) varies as 7/6th power of breakdown voltage. To support the voltage across the P$^+$ substrate/N$^-$ base of an IGBT, the N$^-$-base doping concentration was 1×10^{14} cm^{-3} and thickness = 30 μm. For the same breakdown voltage, what thickness of N$^-$ base will be necessary if a doping concentration = 5×10^{13} cm^{-3} is used?

It is given that

$$V_B \propto N_D^{-3/4} \quad \text{or} \quad V_B = k_1 N_D^{-3/4} \tag{E4.3.1}$$

Also,

$$V_B \propto W_d^{6/7} \quad \text{or} \quad V_B = k_2 W_d^{6/7} \tag{E4.3.2}$$

where k_1 and k_2 are constants. From these equations,

$$k_2 W_d^{6/7} = k_1 N_D^{-3/4} \quad \text{or} \quad W_d = (k_1/k_2)^{7/6} N_D^{-7/8} = k^{7/6} N_D^{-7/8} \tag{E4.3.3}$$

Since, for $N_D = 1 \times 10^{14}$ cm^{-3}, $W_d = 30$ μm $= 30 \times 10^{-4}$ cm, $k^{7/6} = \{W_d/N_D^{-7/8}\} = \{30 \times 10^{-4}/(1 \times 10^{14})^{-7/8}\} = 5.33 \times 10^9$ from which we have for $N_D = 5 \times 10^{13}$ cm^{-3}, $W_d = k^{7/6} N_D^{-7/8} = 5.33 \times 10^9 \times (5 \times 10^{13})^{-7/8} = 5.497 \times 10^{-3}$ cm $\cong 55$ μm.

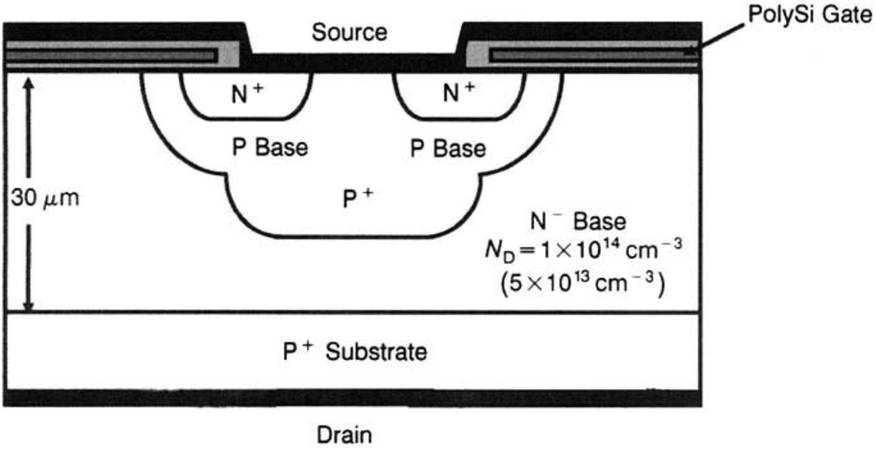

PolySi Gate
Source
N+
N+
P Base
P Base
P+
30 μm
N− Base
$N_D = 1 \times 10^{14}$ cm^{-3}
(5×10^{13} cm^{-3})
P+ Substrate
Drain

Figure E4.4.1 IGBT for Example 4.4.

Example 4.4 The doping concentration of the P^+ substrate in an NPT-IGBT is 5×10^{19} cm^{-3} while the N-base doping density is 1×10^{14} cm^{-3}. Find the depletion region width across the P^+-substrate/N^- base without applying any voltage.

Figure E4.4.1 shows the NPT-IGBT. Barrier potential across the P^+-substrate/N^--base junction is $\phi_0 = (kT/q)\ln(N_A N_D/n_i^2) = 0.0259\ln\{(5 \times 10^{19} \times 1 \times 10^{14})/(1.45 \times 10^{10})^2\} = 0.7977$ V. $x_n = [\{2\varepsilon_0\varepsilon_s(\phi_0 - v_a)N_A\}/qN_D(N_A + N_D)\}]^{1/2} = [\{2 \times 8.854 \times 10^{-14} \times 11.9 \times 0.7977 \times 5 \times 10^{19}\}/\{1.6 \times 10^{-19} \times 1 \times 10^{14} \times (5 \times 10^{19} + 1 \times 10^{14})\}]^{1/2} = 3.24 \times 10^{-4}$ cm $= 3.24$ μm. $x_p = [\{2\varepsilon_0\varepsilon_s(\phi_0 - v_a)N_D\}/\{qN_A(N_A + N_D)\}]^{1/2} = [\{2 \times 8.854 \times 10^{-14} \times 11.9 \times 0.7977 \times 1 \times 10^{14}\}/[1.6 \times 10^{-19} \times 5 \times 10^{19} \times (5 \times 10^{19} + 1 \times 10^{14})\}]^{1/2} = 6.48 \times 10^{-10}$ cm $=$ 6.5×10^{-6} μm $\cong 0$ μm. $\therefore x_d = x_n + x_p = 3.24$ μm $+ 0 = 3.24$ μm.

4.1.4 Current–Voltage (i_d–v_a) Equation

To derive the current–voltage characteristics of a PN junction, we examine the minority-carrier concentrations for the forward-biased diode shown in Fig. 4.4 [4, 5]. Application of forward bias v_a causes minority-carrier injection across the junction where they recombine with majority carriers. The shaded regions indicate the excess minority-carrier concentrations on each side of the junction, starting at the maximum values at $x = 0$ and $x' = 0$, and decreasing to the equilibrium value at large values of x and x'. The excess

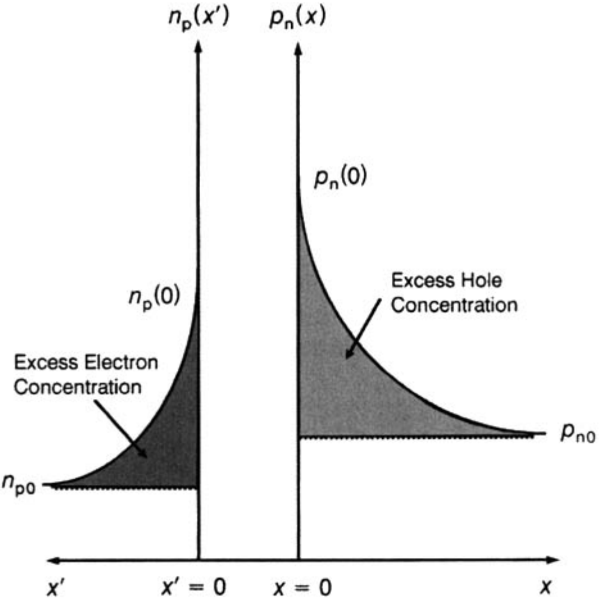


Figure 4.4 Variation of minority carrier concentrations across a P-N junction under forward-biased condition.

hole concentration on the N side at $x = 0$ is written as [5]

$$p_{\mathrm{n}}(0) = p_{\mathrm{n}0} \exp\left(\frac{v_{\mathrm{a}}}{\phi}\right) \tag{4.24}$$

where $p_{\mathrm{n}0} = n_{\mathrm{i}}^2/N_{\mathrm{D}}$ is the equilibrium minority-carrier concentration in N-side and ϕ is the thermal voltage $= kT/q$. This equation is obtained by considering the second and third terms of Eq. (4.3) and replacing $D_{\mathrm{p}}/\mu_{\mathrm{p}}$ by kT/q. Integrating the modified equation $dp/p = -dV/\phi$ across the junction, we get

$$\int_{p_{n0}}^{p_{\mathrm{n}}(0)} \frac{dp}{p} = \int_0^{v_{\mathrm{a}}} \frac{dV}{\phi} \quad \text{or} \quad [\ln p]_{p_{\mathrm{n}0}}^{p_{\mathrm{n}}(0)} = \frac{1}{\phi}[V]_0^{v_a} \quad \text{or}$$

$$\ln p_{\mathrm{n}}(0) - \ln p_{\mathrm{n}0} = \frac{v_{\mathrm{a}} - 0}{\phi} \quad \text{or} \quad \ln\left\{\frac{p_{\mathrm{n}}(0)}{p_{\mathrm{n}0}}\right\} = \frac{v_{\mathrm{a}}}{\phi} \quad \text{or}$$

$$\frac{p_{\mathrm{n}}(0)}{p_{\mathrm{n}0}} = \exp\left(\frac{v_{\mathrm{a}}}{\phi}\right) \quad \text{or} \quad p_{\mathrm{n}}(0) = p_{\mathrm{n}0} \exp\left(\frac{v_{\mathrm{a}}}{\phi}\right).$$

This equation is known as the *law of the junction*. It states that for a forward bias $v_{\mathrm{a}} > 0$, with $v_{\mathrm{a}} \gg \phi$, the hole concentration $p_{\mathrm{n}}(0)$ at the junction in the N side is greatly increased over the thermal equilibrium concentration $p_{\mathrm{n}0}$. In the same manner, excess electron concentration on the P side at $x' = 0$ is written as [5]

$$n_{\mathrm{p}}(0) = n_{\mathrm{p}0} \exp\left(\frac{v_{\mathrm{a}}}{\phi}\right) \tag{4.25}$$

where $n_{\mathrm{p}0} = n_{\mathrm{i}}^2/N_{\mathrm{A}}$ is the equilibrium minority-carrier concentration in P side.

The current flowing in the PN junction varies as the gradient of the excess minority-carrier concentration at $x = 0$ or $x' = 0$. The hole current density flowing in N-type material is expressed by the diffusion equation as (Appendix 4.1)

$$J_{\mathrm{p}}(x) = -qD_{\mathrm{p}} \left.\frac{dp_{\mathrm{n}}}{dx}\right|_{x=0} \tag{4.26}$$

where D_{p} is the diffusion coefficient of holes in N-type semiconductor. Defining the excess hole concentration in N-type material as

$$p_{\mathrm{n}}'(x) = p_{\mathrm{n}}(x) - p_{\mathrm{n}0} \tag{4.27}$$

we obtain from the solution of the continuity equation (Appendix 4.3)

$$\frac{\partial^2 p_n}{\partial x^2} - \frac{p_n - p_{n0}}{L_p{}^2} = 0 \tag{4.28}$$

Using the boundary condition of Eq. (4.28) given by Eq. (4.24) and $p_n|_{x=\infty} = p_{n0}$, we find that the decay of minority carriers away from the junction takes place exponentially obeying the equation

$$p_n'(x) = p_n'(0)\exp\left(-\frac{x}{L_p}\right) = \{p_n(0) - p_{n0}\}\exp\left(-\frac{x}{L_p}\right) \tag{4.29}$$

where L_p is the diffusion length of holes in N-type material. By substituting for $p_n(0)$ from Eq. (4.24) into Eq. (4.29), we get

$$p_n'(x) = p_{n0}\left\{\exp\left(\frac{v_a}{\phi}\right) - 1\right\}\exp\left(-\frac{x}{L_p}\right) \tag{4.30}$$

Now the current density due to excess hole concentration in N side is obtained from Eq. (4.26) as

$$\begin{aligned} J_p(0) &= -qD_p \frac{dp_n}{dx}\bigg|_{x=0} \\ &= -qD_p \times \left[p_{n0}\left\{\exp\left(\frac{v_a}{\phi}\right) - 1\right\}\exp\left(-\frac{x}{L_p}\right) \times -\frac{1}{L_p}\right]_{x=0} \\ &= \frac{qD_p p_{n0}}{L_p}\left\{\exp\left(\frac{v_a}{\phi}\right) - 1\right\} \end{aligned} \tag{4.31}$$

In a similar fashion, the current density due to excess electron concentration in P side is written as

$$J_n(0) = \frac{qD_n n_{p0}}{L_n}\left\{\exp\left(\frac{v_a}{\phi}\right) - 1\right\} \tag{4.32}$$

Assuming that the recombination in the space-charge region is negligibly small, the total current density of the PN junction is the sum of $J_p(0)$ and $J_n(0)$. Then multiplying by the junction area A, the total current flowing in the diode is given by

$$i_D = I_s\left\{\exp\left(\frac{v_a}{\phi}\right) - 1\right\} \tag{4.33}$$

where

$$I_s = qA\left(\frac{D_p p_{n0}}{L_p} + \frac{D_n n_{p0}}{L_n}\right) \tag{4.34}$$

is a constant termed the *saturation current* of the diode. Equation (4.33) is the celebrated *Shockley equation* describing the volt-ampere characteristic of a PN junction diode.

To find the reverse leakage current of the diode, we note that the current due to generation in the depletion region is given by

$$I_{gen} = \frac{qAn_iW}{\tau_{sc}} \tag{4.35}$$

where W is the depletion layer width and τ_{sc} is the space-charge generation lifetime. Total reverse current is the sum of diffusion components in the neutral region and generation current in the depletion region, written as

$$I_R = I_s + I_{gen} = qA\left\{\left(\frac{D_p p_{n0}}{L_p} + \frac{D_n n_{p0}}{L_n}\right) + \frac{n_iW}{\tau_{sc}}\right\} \tag{4.36}$$

Example 4.5 The dopant density of the N$^-$ base in an IGBT is 5×10^{13} cm^{-3}. The recombination lifetime (τ_p) and space-charge generation lifetime (τ_{sc}) are measured to be 10 μsec and 500 μsec, respectively. The chip area is 5 mm × 5 mm. Calculate the reverse leakage current of the P$^+$-substrate/N$^-$-base diode at 1500 V at three different temperatures: (a) 300 K, (b) 400 K, and (c) 500 K. The formula for thermal variation of intrinsic carrier concentration is $n_i = 3.87 \times 10^{16} T^{1.5} \exp(-7.02 \times 10^3)/T$, where T is the absolute temperature. Hole mobility varies with temperature as $495(T/300)^{-2.2}$. Carrier lifetime obeys the relation $\tau(T) = \tau(300)(T/300)^{1.5}$.

The IGBT is shown in Fig. E4.5.1. For a P$^+$N diode, the leakage current equation (4.36) becomes $I_R = qA\ (Wn_i/\tau_{sc} + D_p n_i^2/L_p N_D)$.The temperature-dependent parameters are the intrinsic carrier density n_i, hole diffusion coefficient D_p, space-charge

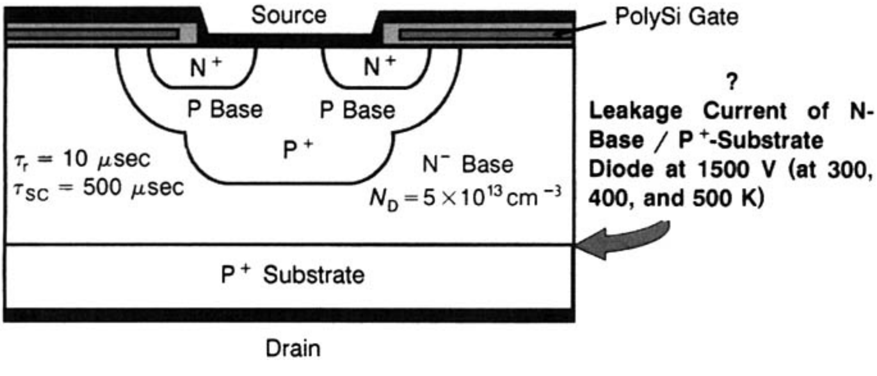


Figure E4.5.1 IGBT for Example 4.5.

Table E4.5.1 Temperature-Sensitive Parameters of Leakage Current

Serial No.	Temperature (K)	n_i (cm^{-3})	μ_p (cm^2/V-sec)	kT/q (V)	D_p (cm^2/sec)	τ_{SC} (μsec)	τ_p (μsec)	L_p (cm)
1	300	1.38×10^{10}	495	0.02588	12.81	500	10	1.13×10^{-2}
2	400	7.4×10^{12}	262.87	0.0345	9.07	769.8	15.4	1.18×10^{-2}
3	500	3.46×10^{14}	160.89	0.0431	6.93	1075.8	21.5	1.22×10^{-2}

Table E4.5.2 Leakage Currents at Different Temperatures

Serial No.	Temperature (K)	I_R (A)
1	300	2.21×10^{-8}
2	400	4.13×10^{-5}
3	500	0.055

generation lifetime τ_{sc}, and hole diffusion length L_p. Using the formulae for thermal variation of various parameters, Einstein's relation $D_p = \mu_p(kT/q)$ and diffusion length $L_p = \surd(D_p\tau_p)$, the calculated values of the parameters at different temperatures are given in Table E4.5.1.

At 1500 V, the depletion layer width $W = \surd\{(2\varepsilon_0\varepsilon_s V_a/(qN_D))\} = \surd\{(2 \times 8.854 \times 10^{-14} \times 11.9 \times 1500/(1.6 \times 10^{-19} \times 5 \times 10^{13}))\} = 198.77 \times 10^{-4}$ cm. On substituting the above values in the I_R equation, the obtained leakage current values are compiled in Table E4.5.2.

4.1.5 Reverse Recovery Characteristics

To study the turn-off switching transition of a power diode D, a DC source V is connected in series with a switch S (Fig. 4.5) [6]. Assuming that S has been ON for a long time, a current I is established in the resistor–inductor (R–L) circuit. Now, if the switch S is turned off, the current in the R–L circuit does not immediately fall to zero due to energy storage in the inductance L. As soon as the current tends to decay, the induced EMF in the inductance forward biases the diode and the current freewheels—that is, continues to flow in the R–L circuit. For large values of L, in concert with rapid toggling of S, the current I remains approximately constant.

When the switch is turned ON again at the instant $t = t_0$, current builds up in the circuit comprising the DC source V and the switch S. The rate of increase of i_s (i.e., di_s/dt) is determined by the residual inductance in the circuit loop consisting of V, D, and S. Constancy of the current I imposes the constraint that the increase of i_s = the decrease of i_d so that $i_s + i_d = I$. Hence, $di_d/dt = -di_s/dt$. At the instant $t = t_1$, the diode current decreases to zero. But the diode continues to conduct in the reverse direction because of the presence of excess minority carriers, which are pushed back across the P-N junction. At $t = t_2$, the excess minority carriers are removed. Now the

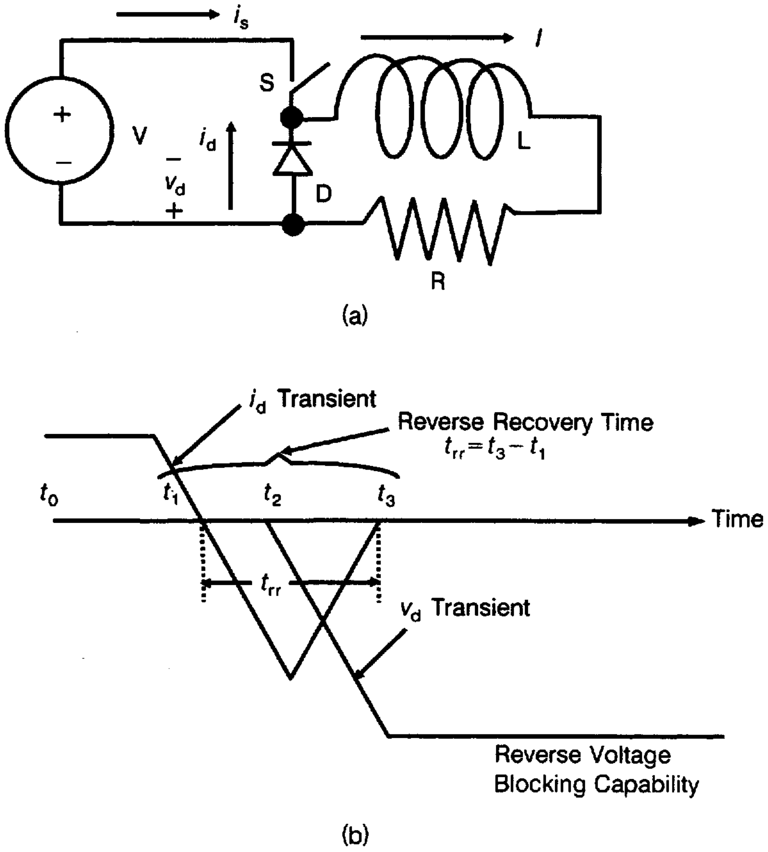

i_s
I
S
V
+
−
i_d
−
v_d
+
L
D
R
(a)
i_d Transient
Reverse Recovery Time
$t_{rr} = t_3 - t_1$
t_0
t_1
t_2
t_3
Time
t_{rr}
v_d Transient
Reverse Voltage
Blocking Capability
(b)

Figure 4.5 Switching of a power diode: (a) Circuit for studying the switching behavior, and (b) voltage and current waveforms for turn-off transition of a power diode.

reverse current begins falling toward zero and the diode recovers its reverse voltage blocking capability. The current flowing from t_2 to t_3 may be viewed as a charging current for the capacitance formed by the diode junction. At $t = t_3$, the diode junction capacitance is charged to the full reverse voltage and the turn-off switching of the diode is completed. The reverse recovery time t_{rr} = the time interval measured from the instant t_1 at which the forward current becomes zero to the instant t_3 at which reverse voltage recovery is completed.

4.2 P-I-N RECTIFIER

The PIN rectifier comprises a high-resistivity or low-doped N-region called the *intrinsic* or *I-region*, sandwiched between P^+ and N^+ end regions [5]. During the forward conduction of the P^+-I-N^+ rectifier (Fig. 4.6), the

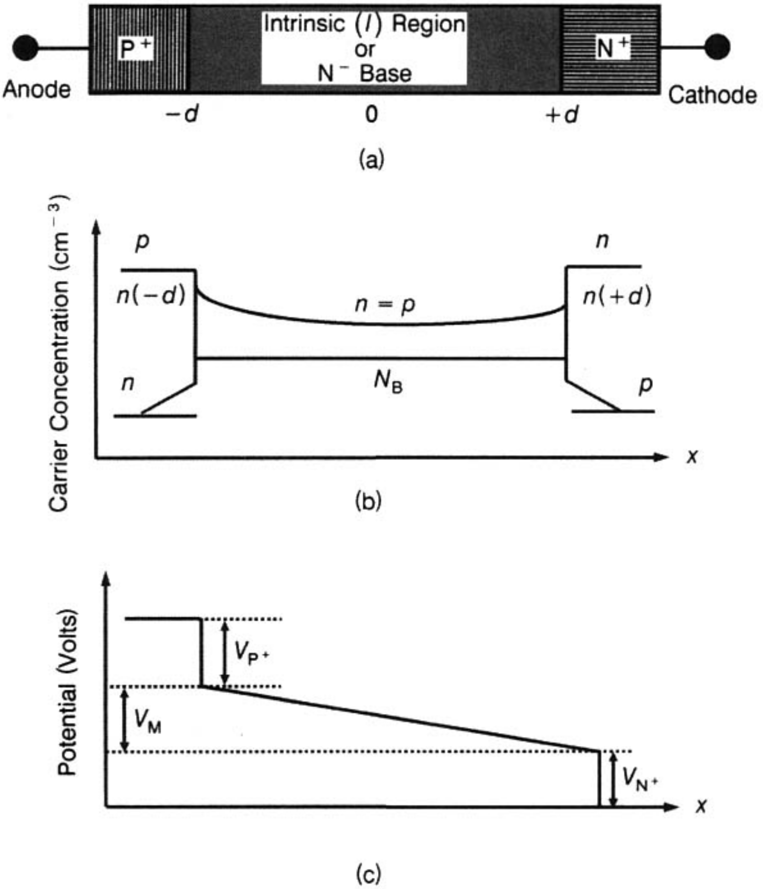

P+
Intrinsic (I) Region or N⁻ Base
N+
Anode
Cathode
−d
0
+d
(a)
Carrier Concentration (cm⁻³)
p
n
n(−d)
n = p
n(+d)
n
N_B
p
x
(b)
Potential (Volts)
V_P+
V_M
V_N+
x
(c)

Figure 4.6 PIN rectifier: (a) Physical structure, (b) carrier distribution profile under high-level injection, and (c) potential distribution at high injection.

intrinsic (I) region is flooded with minority-carrier holes. Under high-level injection, the injected hole concentration is much greater than the background doping density. Conductivity modulation of the intrinsic region lowers the resistance of this region and enables the device to carry a high current during forward conduction. Charge neutrality condition in the N base requires equality of electron and hole concentrations, $n(x) = p(x)$. In the steady state, the current flow takes place through recombination of holes and electrons in the N$^-$ base and the anode–cathode end regions. Assuming negligible recombination in the end regions, the current density depends only on recombination in the N$^-$-base region. It is written as

$$J = \int_{-d}^{+d} qR\,dx = \int_{-d}^{+d} q\frac{n(x)}{\tau_{\mathrm{HL}}}\,dx \tag{4.37}$$

where R is the recombination rate, d is half-width of N^--base region, and τ_{HL} is the high-level lifetime. Taking τ_{HL} to be constant for varying carrier densities, we obtain

$$J = q\frac{n(x)}{\tau_{HL}}\int_{-d}^{+d} dx = q\frac{n}{\tau_{HL}}[x]_{-d}^{+d} = q\frac{n}{\tau_{HL}}[+d-(-d)] = \frac{2qnd}{\tau_{HL}} \tag{4.38}$$

where n is the average carrier density. Considering the N^- base of PIN rectifier, the continuity equation is

$$\frac{\partial n}{\partial t} = 0 = \frac{n}{\tau_{HL}} + D_a\frac{\partial^2 n}{\partial x^2} \tag{4.39}$$

Since the ambipolar diffusion constant (see Note 1 at the end of this section) is related to the ambipolar diffusion length as

$$L_a = \sqrt{D_a\tau_a} \tag{4.40}$$

we have

$$\frac{d^2n}{dx^2} - \frac{n}{L_a^2} = 0 \tag{4.41}$$

Solution of this equation (Appendix 4.4) requires knowledge of boundary conditions. This is obtained from current transport at P^+ and N^+ ends of the diode. At $x = +d$,

$$J_p(+d) = q\mu_p p(+d)\xi(+d) - qD_p\left.\frac{dp}{dx}\right|_{x=+d} = 0 \tag{4.42}$$

Because

$$\xi(+d) = \frac{kT}{q}\frac{1}{n(+d)}\left.\frac{dn}{dx}\right|_{x=+d} \tag{4.43}$$

at $x = +d$, we have

$$J = J_n(+d) = 2qD_n\left.\frac{dn}{dx}\right|_{x=+d} \tag{4.44}$$

Similarly, at $x = -d$,

$$J = J_p(-d) = -2qD_p\left.\frac{dn}{dx}\right|_{x=-d} \tag{4.45}$$

using the charge neutrality condition $n(x) = p(x)$. For the derived boundary conditions, the solution of the continuity equation (4. 41) is expressed as (Appendix 4.4)

$$n = p = \frac{J\tau_{\mathrm{HL}}}{2qL_{\mathrm{a}}}\left\{\frac{\cosh\left(\dfrac{x}{L_{\mathrm{a}}}\right)}{\sinh\left(\dfrac{d}{L_{\mathrm{s}}}\right)} - \frac{1}{2}\frac{\sinh\left(\dfrac{x}{L_{\mathrm{a}}}\right)}{\cosh\left(\dfrac{d}{L_{\mathrm{a}}}\right)}\right\} \tag{4.46}$$

Figure 4.6b shows the carrier distribution in the P-I-N rectifier structure. The electron concentration is highest at the junction of N with $\mathrm{N}^{+}(+d)$, and the hole concentration is highest at the P^{+}-N$(-d)$ junction. The minimum carrier concentration occurs near the cathode side due to different electron and hole mobilities.

The voltage drop is found by estimating the electric field distribution. The electric field is connected with the current $J = J_{\mathrm{n}} + J_{\mathrm{p}}$ through the relations

$$J_{\mathrm{n}} = q\mu_{\mathrm{n}}\left(n\xi + \frac{kT}{q}\frac{dn}{dx}\right) \tag{4.47}$$

and

$$J_{\mathrm{p}} = q\mu_{\mathrm{p}}\left(p\xi - \frac{kT}{q}\frac{dn}{dx}\right) \tag{4.48}$$

from which

$$\xi = \frac{J}{q(\mu_{\mathrm{n}} + \mu_{\mathrm{p}})n} - \frac{kT}{2q}\frac{1}{n}\frac{dn}{dx} \tag{4.49}$$

remembering that charge neutrality must be obeyed. The first term on the right-hand side of this equation represents ohmic voltage drop due to the current flowing through the diode. The second term arises from the asymmetric concentration gradient due to mobility difference between electrons and holes. Then the integration of the electric field distribution ξ given by Eq. (4.49), using the carrier distribution, Eq. (4.46), yields the voltage drop across the middle N base as (Appendix 4.5)

$$\begin{aligned} V_{\mathrm{N}^-} = {} & \frac{3}{2}\left(\frac{kT}{q}\right)\frac{\sinh(d/L_{\mathrm{a}})}{\sqrt{1-(1/4)\tanh^2(d/L_{\mathrm{a}})}}\arctan \\ & \times\left\{\sinh(d/L_{\mathrm{a}})\sqrt{1-(1/4)\tanh^2(d/L_{\mathrm{a}})}\right\} \\ & + \frac{1}{2}\left(\frac{kT}{q}\right)\ln\left\{\frac{1+(1/2)\tanh^2(d/L_{\mathrm{a}})}{1-(1/2)\tanh^2(d/L_{\mathrm{a}})}\right\} \end{aligned} \tag{4.50}$$

Independence of V_{N^-} from current density J originates due to the increase of carrier concentration with current density, nullifying the effect of J. However, there is a drastic increase in V_{N^-} as the ratio d/L_a becomes high. This is because the diffusion length becomes less than the half base width d so that conductivity modulation decreases.

Now let us focus our attention on the junction between P$^+$ anode and N$^-$ base. If p_0 is the thermal-equilibrium minority carrier density in the N$^-$ base and $V_P{}^+$ is the potential drop across P$^+$-N$^-$ junction, we may write

$$p(-d) \approx p_0 \exp\left(\frac{qV_{P+}}{kT}\right) \tag{4.51}$$

from which

$$V_{P+} = \frac{kT}{q} \ln \frac{p(-d)N_d}{n_i^2} \tag{4.52}$$

Similarly,

$$V_{N+} = \frac{kT}{q} \ln \left\{\frac{n(+d)}{N_d}\right\} \tag{4.53}$$

Adding together V_{P+} and V_{N+} we get

$$V_{P+} + V_{N+} = \frac{kT}{q} \ln \left\{\frac{n(+d)n(-d)}{n_i^2}\right\} \tag{4.54}$$

where charge-neutrality condition has been used. Finally, combining ($V_{P+} + V_{N+}$) equation with V_{N^-} and the carrier distribution equations, we can express the current density of a forward-biased P-I-N diode at a given applied voltage under the assumption of negligible recombination in the end regions and Auger recombination (see Note 2 at the end of this section) by the expression (Appendix 4.6)

$$J = \frac{2qD_a n_i}{d} F\left(\frac{d}{L_a}\right) \exp\left(\frac{qV_a}{2kT}\right) \tag{4.55}$$

where (Appendix 4.6)

$$F\left(\frac{d}{L_a}\right) = \frac{d}{L_a} \tanh\left(\frac{d}{L_a}\right) \frac{\exp\left(-\frac{qV_{N^+}}{2kT}\right)}{\sqrt{1-(1/4)\tanh^4(d/L_a)}} \tag{4.56}$$

The forward drop is low at large values of the function $F(d/L_a)$. The maximum value of $F(d/L_a)$ occurs at $d/L_a = 1$. At low values of $d/L_a < 1$, the junction drops dominate and the forward drop falls with rising d/L_a value. At high values of $d/L_a > 1$, the middle region drop dominates and the forward drop increases as the ratio d/L_a becomes larger.

Note 1: Ambipolar Diffusion Constant and Ambipolar Mobility

At high injection levels, electron and hole transport cannot be treated separately because the electrons move in a cloud of holes and conversely. A convenient approach to handle this situation involves combination of continuity equation for electrons with that for holes, and introduction of two new parameters, namely, *ambipolar mobility* μ_a and *ambipolar diffusion constant* D_a, as algebraic functions of electron and hole mobilities and diffusivities. The defining equations for the ambipolar parameters are

$$\mu_a = \frac{\mu_n \mu_p (p-n)}{n\mu_n + p\mu_p} \tag{N1.1}$$

$$D_a = \frac{D_n D_p (p+n)}{nD_n + pD_p} \tag{N1.2}$$

where n and p are the total electron and hole concentrations, μ_n and μ_p are the electron and hole mobilities, and D_n and D_p are the corresponding diffusivities. The advantage gained by this approach is that the resulting single equation allows focusing our attention on the minority carrier concentration, while the presence of majority carrier concentration is automatically accounted for by the ambipolar parameters.

Note 2: Auger Recombination

This recombination mechanism entails the interaction of three free carriers—either two electrons with one hole, or two holes with one electron. Two of these carriers participate in the recombination process while the third carrier takes away the momentum of the incoming carriers as well as the energy released during recombination. The necessity of simultaneous presence of three carriers increases the likelihood of Auger recombination in heavily doped semiconductors. The associated lifetime called the *Auger recombination lifetime* (τ_A) decreases as the square of the dopant concentration. Auger recombination mechanism is the inverse of the impact ionization which causes avalanche breakdown. In impact ionization, a carrier accelerated by the electric field, gives rise to one or more carriers.

Example 4.6 In a PT-IGBT, half-width of N base is 40 μm. The original carrier lifetime is lowered from 10 μsec to 1 μsec and then to 0.5 μsec by electron irradiation. What was the voltage drop across the middle N^--base region initially and after each step? Ambipolar diffusion coefficient is 18.34 cm^2/sec.

The PT-IGBT is shown in Fig. E4.6.1. Initially, the diffusion length was $L_a = \surd(D_a\tau_a) = \surd(18.34 \times 10 \times 10^{-6}) = 0.01354$ cm. So, the original value of d/L_a ratio was $40 \times 10^{-4}/0.01354 = 0.2954$. For this d/L_a ratio, the voltage drop across the middle N^--base region is obtained from Eq. (4.50) as $V_M = 0.0259 \times$

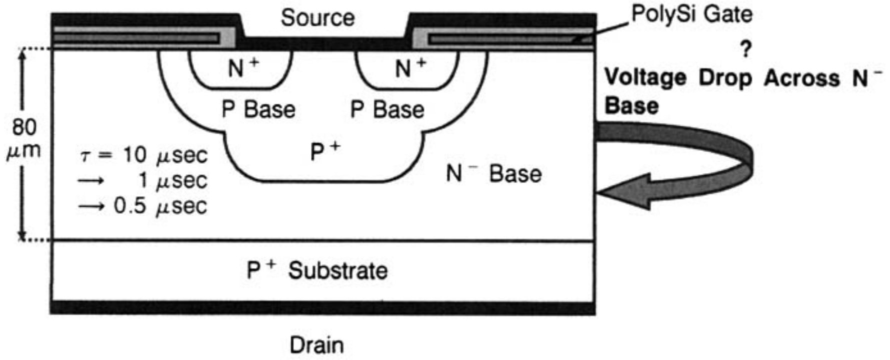

Source
PolySi Gate
?
Voltage Drop Across N⁻ Base
N+
N+
P Base
P Base
P+
80 μm
τ = 10 μsec
→ 1 μsec
→ 0.5 μsec
N⁻ Base
P+ Substrate
Drain

Figure E4.6.1 IGBT for Example 4.6.

[1.5 sinh 0.2954/{1 − 0.25 tanh²(0.2954)}] arctan√{1 − 0.25 tanh²(0.2954) sinh 0.2954} + 0.5 × 0.0259 × ln[{1 + 0.5 tanh²(0.2954)/{1 − 0.5 tanh²(0.2954)}] = 0.0118889 × 44.9113 + 1.068 × 10^{-3} = 0.535 V.

Similarly for $\tau_a = 1$ μsec, $L_a = 4.28 \times 10^{-3}$ cm, d/L_a ratio = 0.9346, V_M = 0.0259 × [1.5 sinh 0.9346/{1 − 0.25 tanh² (0.9346)}] arctan √{1 − 0.25 tanh² (0.9346) sinh 0.9346} + 0.5 × 0.0259 × ln[{1 + 0.5 tanh² (0.9346)/{1 − 0.5 tanh² (0.9346)}] = 0.0483 × 42.766 + 7.127 × 10^{-3} = 2.0656 + 7.127 × 10^{-3} = 2.0727 V.

Likewise for $\tau_a = 0.5$ μsec, $L_a = 3.03 \times 10^{-3}$, d/L_a ratio = 1.32, V_M = 0.0259 × [1.5 sinh 1.32/{1 − 0.25 tanh² (1.32)}] arctan√{1 − 0.25 tanh² (1.32) sinh 1.32} + 0.5 × 0.0259 × ln[{1 + 0.5 tanh² (1.32)/{1 − 0.5 tanh² (1.32)}] = 0.083 × 39.375 + 0.0252775 = 3.268 + 0.0252775 = 3.293 V.

Example 4.7 In the PIN rectifier of an IGBT, half-width of N⁻ base equals the diffusion length (70 μm). If for $d/L_a = 1$, the function $F(d/L_a)$ is 0.25, evaluate the current density flowing at a forward voltage of 500 mV.

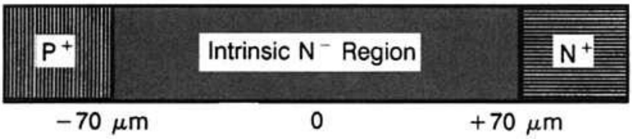

P+
Intrinsic N⁻ Region
N+
−70 μm
0
+70 μm

Figure E4.7.1 PIN rectifier of the IGBT of Example 4.7.

The PIN rectifier of the IGBT is shown in Fig. E4.7.1. Neglecting the recombination in the end regions of the PIN diode, the current density is given by $J = (2qD_a n_i/d)F(d/L_a)\exp\{qV_A/(kT)\}$ = {(2 × 1.6 × 10^{-19} × 18.34 × 1.45 × 10^{10})/(70 × 10^{-4})} × 0.25 × exp(0.2/0.0259) = 3.0392 × 10^{-6} × 2.42 × 10^{8} = 735.91 A/cm².

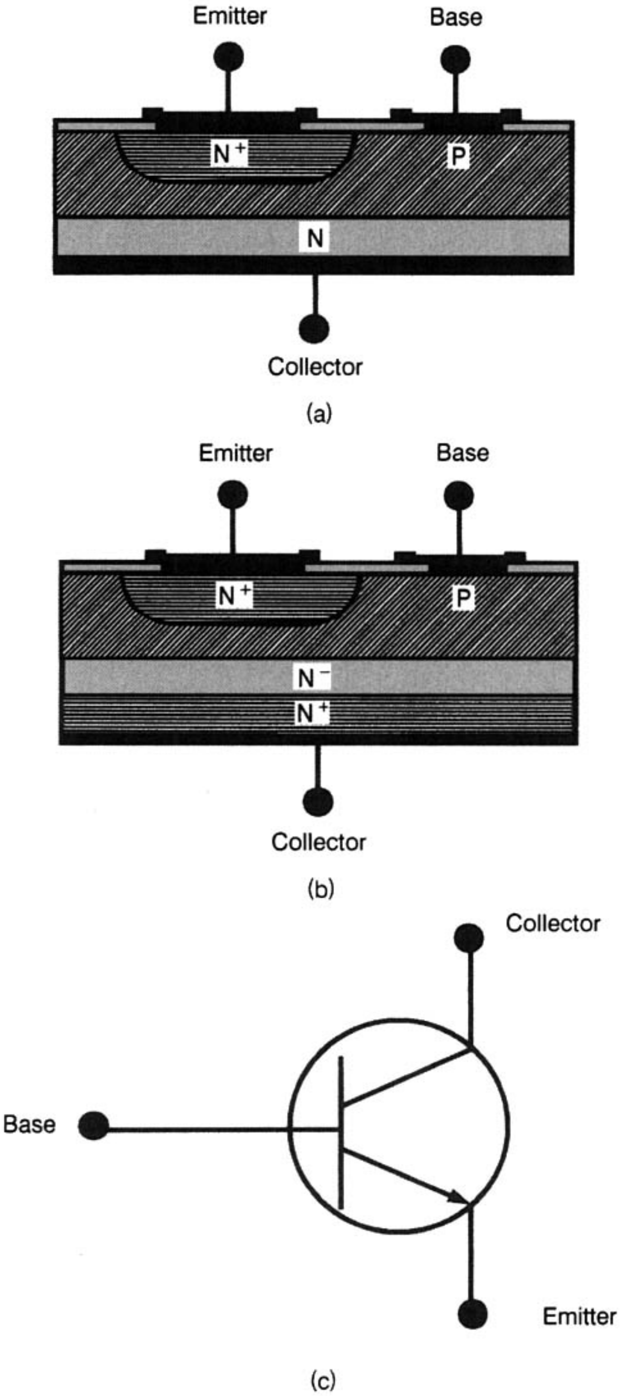

Emitter
Base
N+
P
N
Collector
(a)
Emitter
Base
N+
P
N−
N+
Collector
(b)
Collector
Base
Emitter
(c)

Figure 4.7 NPN bipolar transistor: (a) Double-diffused structure, (b) triple-diffused structure, and (c) circuit diagram symbol.

4.3 BIPOLAR JUNCTION TRANSISTOR

High-power discrete transistors [8–12] with current ratings of several hundred amperes and voltage ratings of several hundred volts are used as static switches in power converters. Both NPN and PNP types are available, but the ratings for NPN variety are higher. Figure 4.7a presents the junction structure of a double-diffused NPN transistor, Figure 4.7b shows the triple-diffused or epitaxial NPN transistor, and Fig. 4.7c gives the circuit diagram symbol of NPN device. In the triple-diffused transistor, an extra N^+ layer is formed on the backside of the N^- wafer to provide a low-resistance ohmic contact between the collector region and the collector metal layer. In the epitaxial transistor, the starting material is an N^+ substrate over which an N^- layer is grown by epitaxy.

4.3.1 Static Characteristics and Current Gain

The static characteristics of the transistor are derived from the P-N junction theory. The continuity and current density equations govern the characteristics. The emitter and collector currents of a P-N-P transistor as a function of applied voltages are given by

$$\begin{aligned} I_{\mathrm{E}} &= qA\left(\frac{D_{\mathrm{p}}p_{\mathrm{B}}}{L_{\mathrm{B}}}\right)\coth\left(\frac{W}{L_{\mathrm{B}}}\right) \\ &\times\left[\left\{\exp\left(\frac{qV_{\mathrm{EB}}}{kT}\right)-1\right\}-\operatorname{sech}\left(\frac{W}{L_{\mathrm{B}}}\right)\left\{\exp\left(\frac{qV_{\mathrm{CB}}}{kT}\right)\text{-}1\right\}\right] \\ &+qA\left[\left(\frac{D_{\mathrm{E}}n_{\mathrm{E}}}{L_{\mathrm{E}}}\right)\left\{\exp\left(\frac{qV_{\mathrm{EB}}}{kT}\right)-1\right\}\right] \end{aligned} \tag{4.57}$$

and

$$\begin{aligned} I_{\mathrm{C}} &= qA\left(\frac{D_{\mathrm{p}}p_{\mathrm{B}}}{L_{\mathrm{B}}}\right)\operatorname{cosech}\left(\frac{W}{L_{\mathrm{B}}}\right) \\ &\times\left[\left\{\exp\left(\frac{qV_{\mathrm{EB}}}{kT}\right)-1\right\}-\cosh\left(\frac{W}{L_{\mathrm{B}}}\right)\left\{\exp\left(\frac{qV_{\mathrm{CB}}}{kT}\right)-1\right\}\right] \\ &-qA\left[\left(\frac{D_{\mathrm{C}}n_{\mathrm{C}}}{kT}\right)\left\{\exp\left(\frac{qV_{\mathrm{CB}}}{kT}\right)-1\right\}\right] \end{aligned} \tag{4.58}$$

where A is the cross-sectional area of the transistor, D_E, D_B, and D_C are the minority-carrier diffusion coefficients: L_E, L_B, and L_C are the minority-carrier diffusion lengths in emitter, base and collector; n_E, n_C are the equilibrium minority-carrier (electron) densities in the emitter and collector; p_B is the equilibrium minority-carrier density in the base; W is the base width, and V_{EB} and V_{CB} are the emitter-base and collector-base voltages. The difference $I_E - I_C = I_B$ appears as the base current.

The common-base current gain

$$\alpha = h_{FB} = \frac{\partial I_C}{\partial I_E} = \left(\frac{\partial I_{pE}}{\partial I_E}\right)\left(\frac{\partial I_{pC}}{\partial I_{pE}}\right)\left(\frac{\partial I_C}{\partial I_{pC}}\right) \tag{4.59}$$

$$= \text{Emitter injection efficiency } (\gamma) \times \text{Base transport factor } (\alpha_T)$$

$$\times \text{ Collector multiplication factor } (M)$$

$$\alpha = \gamma \alpha_T M \tag{4.60}$$

For transistor operation at collector–base voltage ≪ avalanche breakdown voltage,

$$\alpha = \gamma \alpha_T \tag{4.61}$$

The parameter γ is given by

$$\gamma = 1 - \left(\frac{D_E}{D_p}\right)\left(\frac{N_B}{N_E}\right)\left(\frac{W}{L_E}\right) \tag{4.62}$$

where D_E is the minority-carrier (electron) diffusion coefficient in the emitter, D_p is the hole diffusion coefficient, N_B and N_E are the dopant densities of the base and emitter, W is the base width, and L_E is the minority-carrier diffusion length in the emitter. Also [5],

$$\alpha_T = 1 - \frac{W^2}{2{L_B}^2} \tag{4.63}$$

This equation is derived in Appendix 4.8. To improve the injection efficiency γ, emitter doping must be much greater than base doping, but bandgap narrowing and Auger recombination restrict γ.

The static common-emitter current gain β equals $h_{FE} = \partial I_C / \partial I_B$. The current gains α and β are related as follows [5]:

$$\beta = \frac{\alpha}{1 - \alpha} \tag{4.64}$$

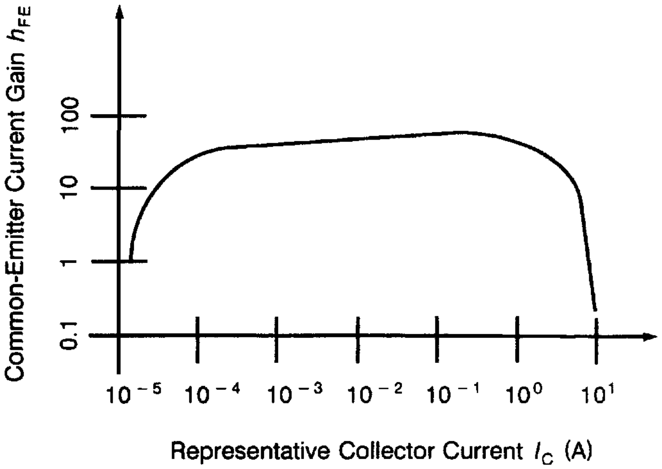


Figure 4.8 Typical variation of common-emitter current gain h_{FE} of a transistor with collector current.

The variation of current gain β with collector current I_C is shown in Fig. 4.8. At low collector current values, the recombination–generation current in the emitter depletion region is small and the surface leakage current is much greater than useful diffusion current of minority carriers across the base. As a result, γ is small and the gain is low. Bulk and surface trap minimization improves β at low current levels. As collector current increases, β rises to a high plateau. For still higher I_C values, the injected minority-carrier density in the base exceeds the majority-carrier density (high-level injection). Effective base concentration increases, lowering the injection efficiency. This conductivity modulation is known as *Webster's effect* [13]. In lightly doped epitaxial transistors, the high-field region is relocated from N^+-emitter/P-base junction to N^--epitaxial layer/N^+-substrate junction of the N^+-emitter/P-base/N^--epitaxial layer/N^+-substrate power transistor structure having epitaxial collector layer. Under high-level injection, the effective base width increases from the P-base width to (P-base width + N-epitaxial layer width). The classical concept of well-defined emitter-base and collector-base transition regions no longer holds. Base width modulation is known as *Kirk's effect* [14]. Base widening results in decrease of β.

The basic model of a bipolar transistor is the *Ebbers–Moll model* [15] consisting of two diodes connected back-to-back and two current sources driven by the diode currents, assumed to have ideal characteristics. The *Gummel–Poon model* [16] is based on an integral charge-control relation relating the terminal electrical characteristics to base charge.

Example 4.8 Find the common-base current gain of the N^+PN transistor in an IGBT having N^+ layer concentration $= 1 \times 10^{19}$ cm^{-3}, P-base concentration $= 2 \times$

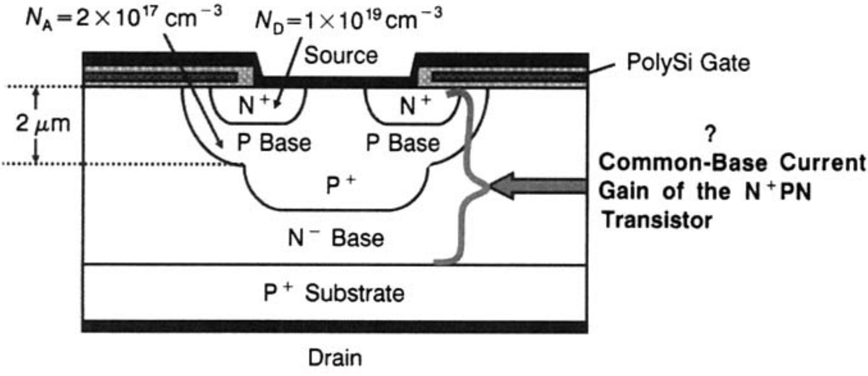

$N_A = 2\times10^{17}$ cm^{-3}
$N_D = 1\times10^{19}$ cm^{-3}
Source
PolySi Gate
2 μm
N+
N+
P Base
P Base
P+
N⁻ Base
P+ Substrate
Drain
?
Common-Base Current Gain of the N+PN Transistor

Figure E4.8 I IGBT for Example 4.8.

10^{17} cm^{-3}, and P-base width = 2 μm, given that hole diffusion coefficient in the emitter = 1 cm^2/sec, hole diffusion length in the emitter = 1 μm, electron diffusion coefficient in the base = 35 cm^2/sec, and electron lifetime in base = 1 μsec.

The IGBT and its structural parameters are shown in Fig. E4.8.1.

$$\gamma = 1 - (1/35)\{(2\times10^{17}\ \text{cm}^{-3})/(1\times10^{19}\ \text{cm}^{-3})\}(2/1) = 0.998857$$

$$\alpha_T = 1 - (2\times10^{-4})^2/\left[2\{\surd(35\times1\times10^{-6})^2\}\right] = 0.9994286$$

$$\alpha = \gamma\alpha_T = 0.998286.$$

Example 4.9 Calculate the common-base current gain of the P$^+$NP transistor in an IGBT having P$^+$-layer concentration = 1×10^{19} cm^{-3}, N$^-$-base concentration = 1×10^{14} cm^{-3}, and N$^-$ base width = 80 μm, given that electron diffusion coefficient in the emitter = 1 cm^2/sec, electron diffusion length in the emitter = 1 μm, hole diffusion coefficient in the base = 12.4 cm^2/sec, and hole lifetime in base = 10 μsec.

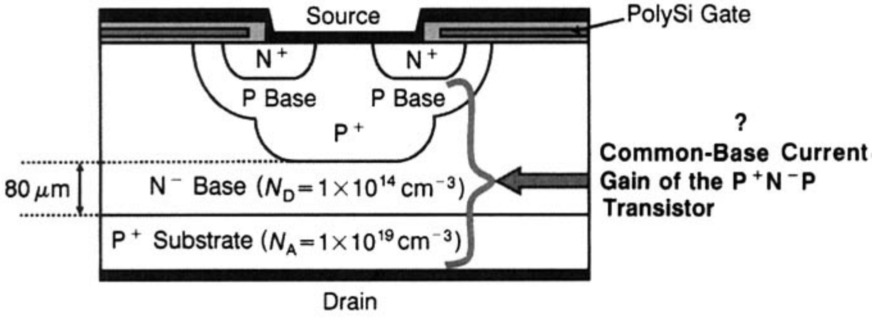

Source
PolySi Gate
N+
N+
P Base
P Base
P+
80 μm
N⁻ Base ($N_D = 1\times10^{14}$ cm^{-3})
P+ Substrate ($N_A = 1\times10^{19}$ cm^{-3})
Drain
?
Common-Base Current Gain of the P+N⁻P Transistor

Figure E4.9.1 IGBT for Example 4.9.

Figure E4.9.1 depicts the IGBT of this problem.

$$\gamma = 1 - (1/12.4)\{(1 \times 10^{17}\,\mathrm{cm}^{-3})/(1 \times 10^{19}\,\mathrm{cm}^{-3})\}(80/1) = 0.93548$$

$$\alpha_T = 1 - (80 \times 10 - 4)^2/\left[2\{\surd(12.4 \times 10 \times 10^{-6})^2\}\right] = 0.7419$$

$$\alpha = \gamma\alpha_T = 0.6941.$$

4.3.1 Power Transistor Switch

Here, Fig. 4.9a, the control input is applied at the base terminal, the control circuit is connected between base and emitter, and the power terminals are collector and emitter with the load resistance R and the DC supply voltage as shown. The plots of collector current I_C versus collector–emitter voltage V_{CE}, for different values of base drive I_B, are shown in Fig. 4.9b. These are the *output characteristics* of the transistor. With reference to these characteristics, the two states of transistor switch are:

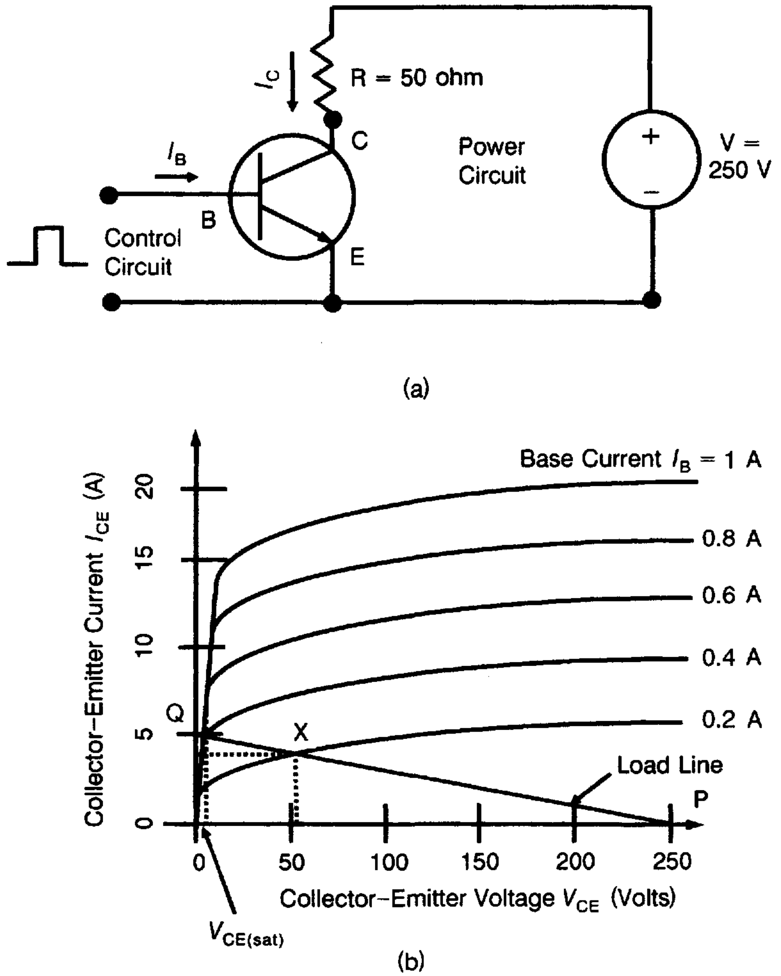
(i) *OFF State or Cut-off Condition*: For base–emitter voltage $V_{BE} < 0$, base current $I_B = 0$ and collector current $I_C = 0$.

(ii) *ON State*: Suppose $V = 250$ V, $R = 50$ ohm and $I_B = 0.4$ A. Applying Kirchhoff's law to the power circuit loop, we have

$$V_{CE} = V - I_C R \tag{4.65}$$

Choosing two points, $I_C = 0$, $V_C = V$ and $V_{CE} = 0$, $I_C = V/R$, gives us a straight line representing the above equation. This straight line PQ is called the *load line*, and its intersection (point X) with the output characteristics ($I_B = 0.4$-A curve) yields the current through the switch (4 A) and the voltage across it (55 V). For $R = 50$ ohm, we find that load line intersects the $I_B = 0.4$-A characteristic on the extreme left where the characteristics for different I_B values are overlapping each other, known as the *saturation region.* In this condition, the voltage drop across the transistor is very small. It is called the *saturation voltage* $V_{CE(sat)}$. But it is readily obvious that for I_B values, below a certain level, $V_{CE(sat)}$ becomes very high, causing excessive power dissipation. For ensuring a saturated ON state, sufficient base drive must be provided. The minimum base current required to furnish a saturated ON state is given by

$$I_B = \frac{I_C}{h_{FE}} \tag{4.66}$$


I_C
R = 50 ohm
I_B
C
B
E
Control Circuit
Power Circuit
+
−
V = 250 V
Collector–Emitter Current I_{CE} (A)
0
5
10
15
20
Base Current I_B = 1 A
0.8 A
0.6 A
0.4 A
0.2 A
Q
X
Load Line
P
0
50
100
150
200
250
Collector–Emitter Voltage V_{CE} (Volts)
$V_{CE(sat)}$

(a)

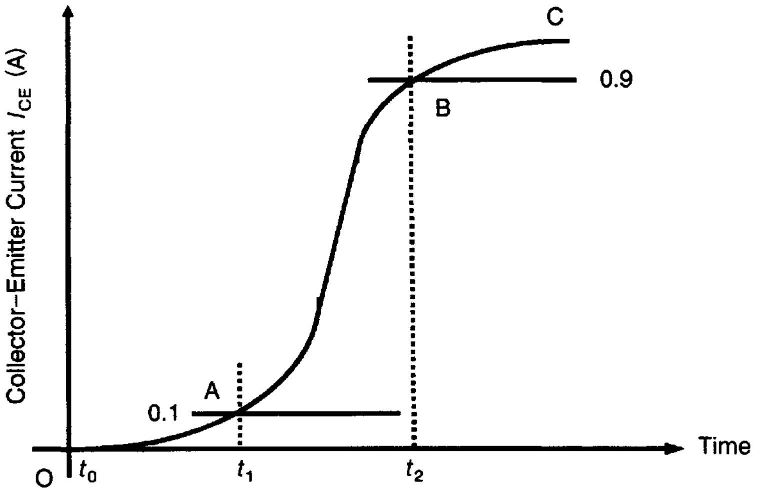
(b)

Figure 4.9 Use of power transistor as a switch: (a) The switching circuit and (b) the output characteristics and the load line.

where h_{FE} is the *common emitter current gain* of the transistor. For adequate safety margin, the circuit is made to operate with a high value of I_B and is said to work with a "forced h_{FE}." Overdriving the base increases the turn-off time due to the excessive minority-carrier injection from the emitter into the base. A "proportional drive" is useful. Herein the base current increases or decreases in accordance with the collector current magnitude.

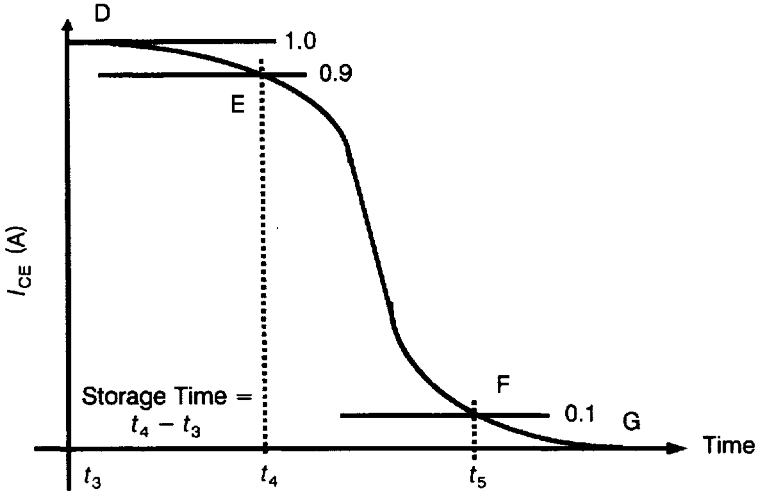
4.3.2 Transistor Switching Times

The time instants marked in the turn-on and turn-off switching transitions (Fig. 4.10) have the following meanings: t_0 (point O) is the instant at which



Rise Time = $t_2 - t_1$

(a)



Fail Time = $t_5 - t_4$

(b)

Figure 4.10 Switching transitions of a transistor: (a) Turn-on transition and (b) turn-off transition.

the base current pulse arrives and the turn-on process starts. t_1 (point A) is the instant at which the collector current rises to 0.1 of its final value, and t_2 (point B) is the instant at which it acquires 0.9 of the final value (point C). t_3 (point D) is the instant at which turn-off is started by applying a small reverse voltage on the base, producing a reverse base current due to excess minority carriers in the base. After t_3, this current persists for a short duration unless the minority carriers have decayed to zero. t_4 (point E) and t_5 (point F) are the instants at which the collector current has fallen to 0.9 and 0.1 of its ON-state value (point G).

With respect to the above time instants, the following time delays are defined: *Rise time* $(t_r) = t_2 - t_1$, *storage time* $(t_s) = t_4 - t_3$, *and fall time* $(t_f) = t_5 - t_4$.

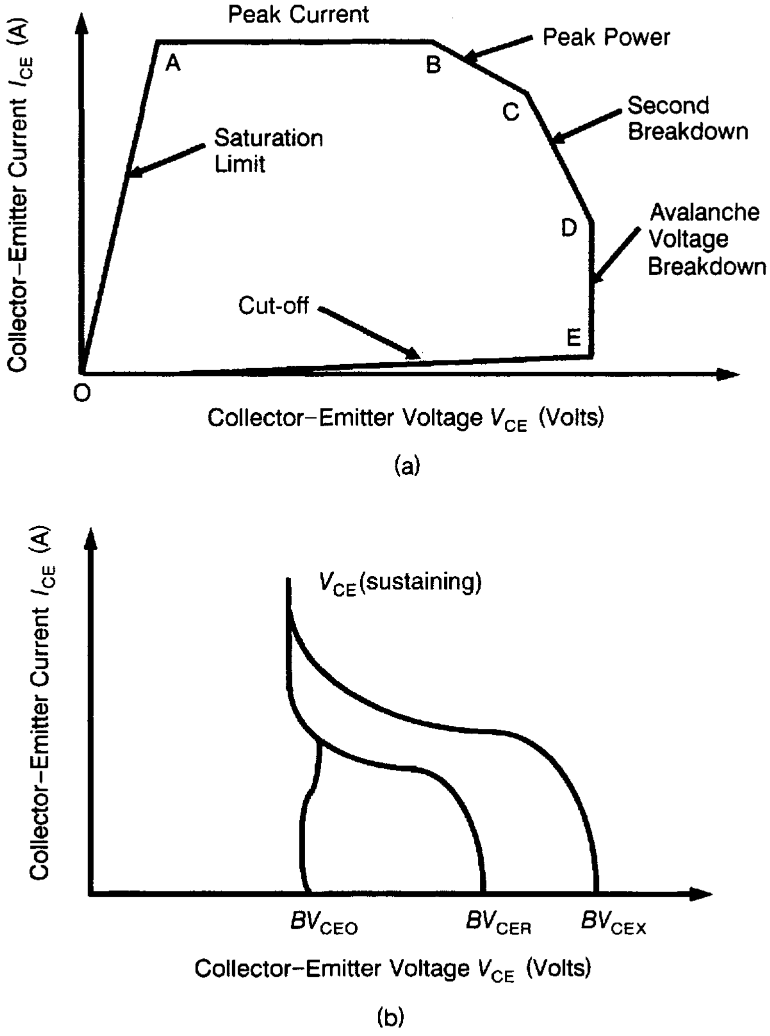


Figure 4.11 (a) Reverse-biased safe operating area (RBSOA) of a power transistor showing the various boundary lines and (b) limits of the avalanche breakdown voltage. **Nomenclature:** $V_{CE(sustaining)}$ = collector-to-emitter sustaining voltage, BV_{CEO} = collector-to-emitter breakdown voltage with base terminal open, BV_{CER} = collector-to-emitter breakdown voltage with impedance R connected between base and emitter, BV_{CEX} = collector-to-emitter breakdown voltage with base terminal shorted to emitter.

4.3.3 Safe Operating Area (SOA)

Transitions between operating points occurring during switching or otherwise must be confined within finite boundaries on the I_C versus V_{CE} plane, called the *safe operating area* (Fig. 4.11a), which has the limits as follows: (i) *Cut-off and Saturation Boundaries* (OE and OA): Normal operation is restricted above the cut-off line and to the right of the saturation line. (ii) *Peak Current Boundary* (AB): This refers to the maximum permissible collector current. (iii) *Peak Power Boundary* (BC): Collector power dissipation $p_{\max} = V_{CE} I_C$.

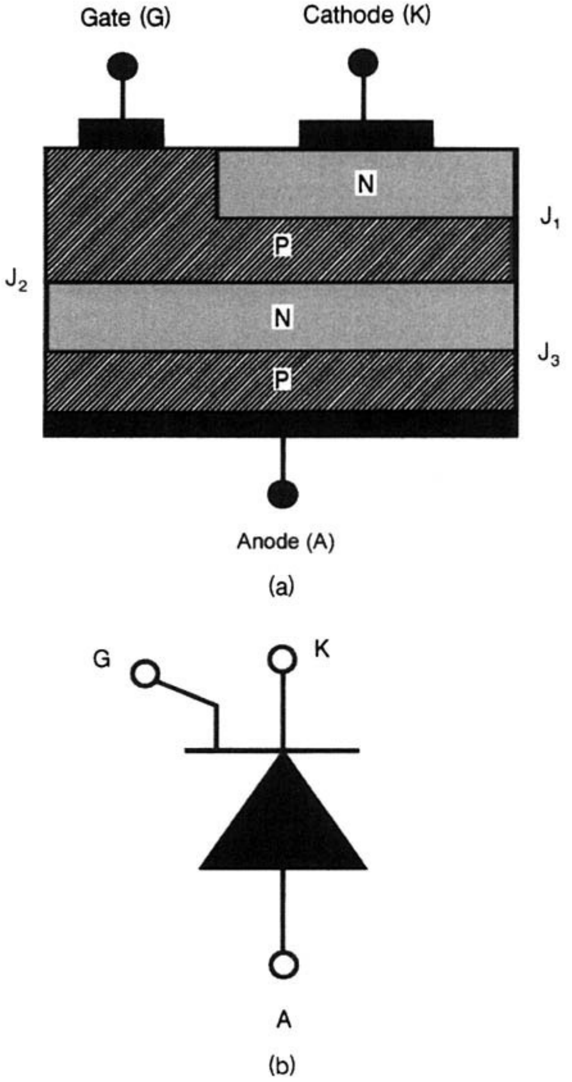


Figure 4.12 Power thyristor: (a) Cross section and (b) circuit symbol.

(iv) *Second Breakdown Boundary* (CD): This phenomenon occurs when voltage, current, and power dissipation, though simultaneously high, are below the above-mentioned levels. For example, nonuniform current distribution during turn-off transition produces local *hot spots*, causing transistor burnout. (v) *Avalanche Voltage Breakdown Boundary* (DE): On increasing V_{CE}, the avalanche breakdown voltage is determined by the manner of connection of the base terminal, Fig.4.11b. Breakdown voltage with base terminal open (BV_{CEO}) is less than breakdown voltage with base shorted to emitter (BV_{CEX}). Breakdown voltage (BV_{CER}) for intermediate values of base-to-emitter resistance (R) lies between BV_{CEO} and BV_{CEX}. In all cases, after avalanche breakdown has taken place, the voltage tends to remain constant at a particular value called the *sustaining voltage* BV_{CES}.

4.4 THYRISTOR

Power thyristors [17, 18] are available in current ratings from a few milliamperes to 5000 A and voltage ratings more than 10,000 V, housed in leaded plastic, stud-type, and disk-type or hockey-puck casings, according to the ratings. Thyristors are classified as *slow thyristors* used for line commutated converters working on low power-line frequencies (50–60 Hz) and *fast thyristors* employed in force commutated converters at high switching frequencies. The thyristor is a four-layer structure (Fig. 4.12) with three internal junctions J_1, J_2, and J_3 connected in series and having three terminals: *anode* (A) with metallic contact on the outer P layer, *cathode* (K) on the outer N layer, and *gate* (G) on the inner P layer. A and K are the power terminals, while the control signal is applied between G and K.

4.4.1 Operating States of Thyristor

The thyristor is a bistable switch with a low-impedance ON state and high-impedance OFF state. Its static characteristics are shown in Fig. 4.13.

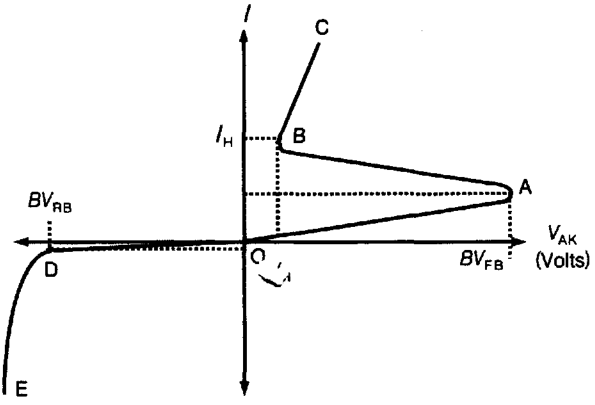


Figure 4.13 Static characteristics of a thyristor.

Region *O-A* of the thyristor is the *forward blocking state*, Region *A-B* is the *negative resistance region*, Region *B-C* is the *forward conducting state*, region *C-D* is the *reverse blocking state*, and region *D-E* is the *reverse breakdown state*. While there is only one ON state, namely, the forward conducting ON state (*B-C*), there are two OFF states, namely, the forward blocking OFF state (*O-A*) and reverse blocking OFF state (*O-D*). The gate can switch the thyristor from the forward blocking OFF state to the forward conducting ON state but cannot implement the reverse transition. Figure 4.14 displays the biasing conditions in the three operating states of the thyristor.

In the *reverse blocking OFF state* (Fig. 4.14a), junctions J_1 and J_3 are reverse-biased and J_2 is forward-biased so that the thyristor cannot conduct except for the small leakage current. In practical thyristors, the reverse blocking voltage is mainly decided by the junction J_1 because the breakdown voltage of J_3 is very small in comparison.

In the *forward blocking OFF state* (Fig. 4.14b), J_2 is reverse-biased while other junctions J_1 and J_3 are forward-biased. Practically, the forward and reverse blocking voltages are identical unless specifically designed otherwise. So the thyristor is a symmetrical voltage blocking device.

Switching of thyristor to the *forward conducting ON state* (Fig. 4.14c) is understood by referring to the two-transistor analogy of the device (Fig. 4.15). Dissecting the thyristor into a combination of a PNP transistor Q_1 and an NPN transistor Q_2 by an imaginary plane, as shown, initially the gate current I_G is zero and both the transistors Q_1, Q_2 are in the nonconducting state. On applying a small gate current, this current serves as the base current I_{B2} of transistor Q_2. Therefore Q_2 starts conducting, resulting in a collector current I_{C2}. But the collector current I_{C2} acts as the base current I_{B1} of transistor Q_1. Due to the base current I_{B1}, a collector current I_{C1} starts to flow. This collector current I_{C1} supplies the base current to transistor Q_2, further increasing I_{C2}. Thus within a few microseconds, a *regenerative mechanism* is initiated driving both the transistors Q_1 and Q_2 into their saturated ON states which are maintained even if the gate current is withdrawn, provided that the current level of thyristor is above a minimum limit called the *holding current*.

To analyze forward switching, we note that the base current I_{B1} of the P-N-P transistor is given by

$$I_{B1} = (1 - \alpha_1) I_A - I_{CO1} \tag{4.67}$$

where α_1 is the common-base current gain of the P-N-P transistor, I_A is the anode current and I_{CO1} is the collector–base reverse saturation current. This base current is supplied by the collector of the N-P-N transistor. Now, the collector current I_{C2} of the N-P-N transistor is expressed as

$$I_{C2} = \alpha_2 I_K + I_{CO2} \tag{4.68}$$

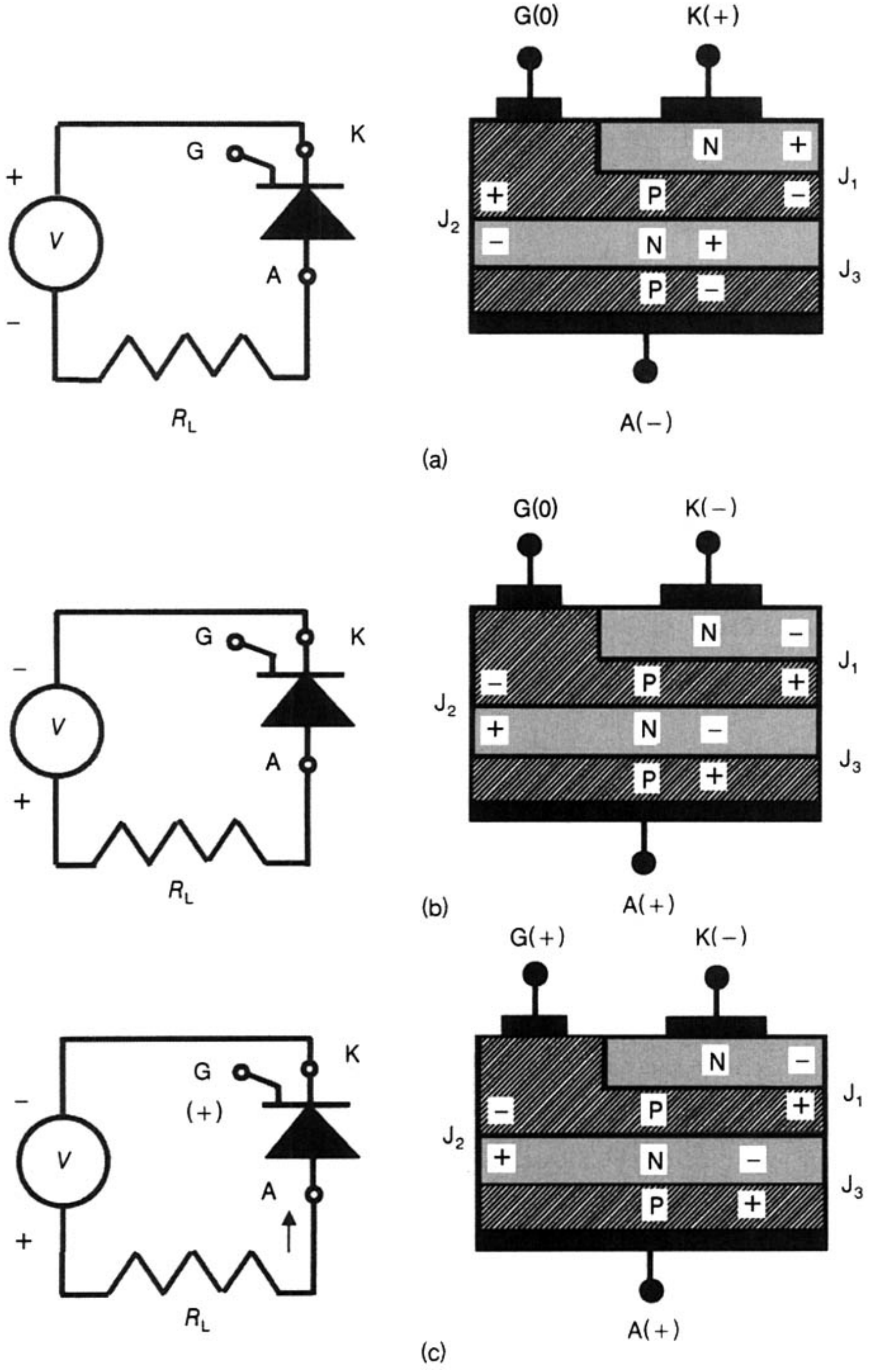

G(0)
K(+)
+
V
−
G
K
A
R_L
N
+
P
+
−
J_1
J_2
−
N
+
P
−
J_3
A(−)
(a)
G(0)
K(−)
−
V
+
G
K
A
R_L
N
−
P
−
+
J_1
J_2
+
N
−
P
+
J_3
A(+)
(b)
G(+)
K(−)
−
V
+
G
(+)
K
A
R_L
N
−
P
−
+
J_1
J_2
+
N
−
P
+
J_3
A(+)
(c)

Figure 4.14 Different operating states of a thyristor: (a) Reverse blocking OFF state, (b) forward blocking OFF state, and (c) forward conducting ON state.

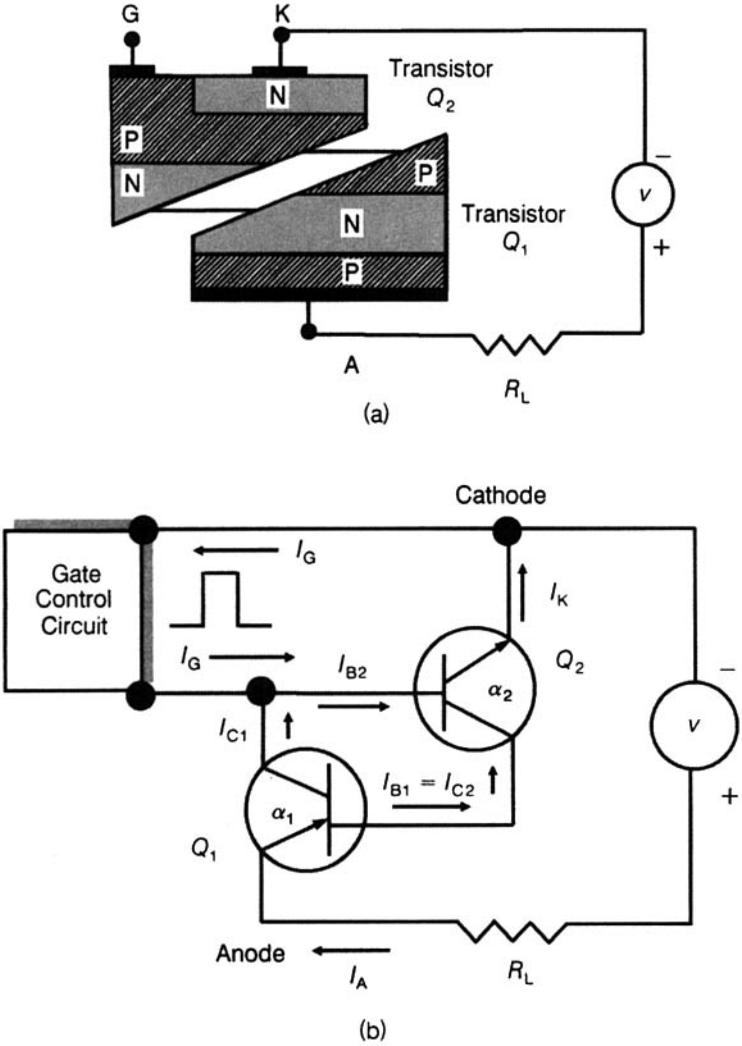

G
K
Transistor Q_2
N
P
N
P
N
P
Transistor Q_1
V
−
+
A
R_L
(a)
Cathode
Gate Control Circuit
I_G
I_G
I_{B2}
I_K
Q_2
α_2
I_{C1}
α_1
$I_{B1} = I_{C2}$
Q_1
V
−
+
Anode
I_A
R_L
(b)

Figure 4.15 Two-transistor equivalent circuit of a thyristor.

where α_2 is the common-base current gain of the N-P-N transistor and I_{CO2} is collector–base reverse saturation current. Since $I_{B1} = I_{C2}$ we have

$$(1 - \alpha_1) I_A - I_{CO1} = \alpha_2 I_K + I_{CO2} \tag{4.69}$$

But $I_K = I_A + I_G$, hence

$$I_A = \frac{\alpha_2 I_G + I_{CO1} + I_{CO2}}{1 - (\alpha_1 + \alpha_2)} \tag{4.70}$$

The current gains α_1 and α_2 are functions of current I_A and rise with increasing current. When $\alpha_1 + \alpha_2 = 0$, $I_A = \infty$, forward breakover occurs and the device behaves as a P-I-N diode.

4.4.2 *di/dt* Capability of Thyristor and Inability of Turn-Off by Reverse Gate Current Pulse

A thyristor T may be imagined as comprising several elementary thyristors $T_1, T_2, T_3, \ldots, T_N$ in parallel. When a gate current pulse is applied, first the thyristor T_1 nearest to the gate electrode is turned on. Part of the current of this thyristor goes to the gate of T_2, causing it to turn on. Likewise, T_2 turns on T_3, and so on up to the Nth thyristor. Thus a finite time elapses before all the constituent thyristors are turned on, making the device fully conducting. Spreading of the current plasma is characterized by a *spreading velocity* [19]. Thus if the current in the external circuit rises faster than the speed at which the switching progresses across the area of thyristor pellet, excessive current density and hence local heating takes place in regions near the gate, resulting in permanent device damage. This imparts to the thyristor a di/dt rating, and this limit should not be exceeded for safe operation. This also explains why the thyristor cannot be switched off through the gate electrode because the gate electrode is away from a large part of the cathode area and is therefore ineffective in controlling the same. By using an *interdigitated gate–cathode geometry* or *involute pattern* [20], the di/dt rating of thyristor is increased as the gate is distributed to lie within a short distance of the cathode all over the pellet. Another method is to use an *amplifying gate* that serves as a pilot device, turning on rapidly due to its small lateral dimensions and sending a strong driving current to the main device. The higher this driving current, the larger the initial turn-on area.

4.4.3 *dv/dt* Rating of Thyristor

When the forward applied voltage rises at a very fast rate dv/dt, the thyristor is turned on even in the absence of a gate current pulse. This is because during the forward blocking state (Fig. 4.14b) the junction J_2 is reverse-biased, behaving as a charged capacitance C. A rapid dv/dt change produces a *capacitive displacement current* Cdv/dt that turns on the thyristor like a gate current pulse if its magnitude exceeds a threshold value. By using a *shorted emitter connection* (Fig. 4.16), in which local short circuits are produced by emitter metal overlap over the base regions of NPN transistor, the thyristor current flows directly to the cathode terminal via an alternative path without contributing to the gate current. This bypassing of current raises the dv/dt threshold for false turn-on switching [21].

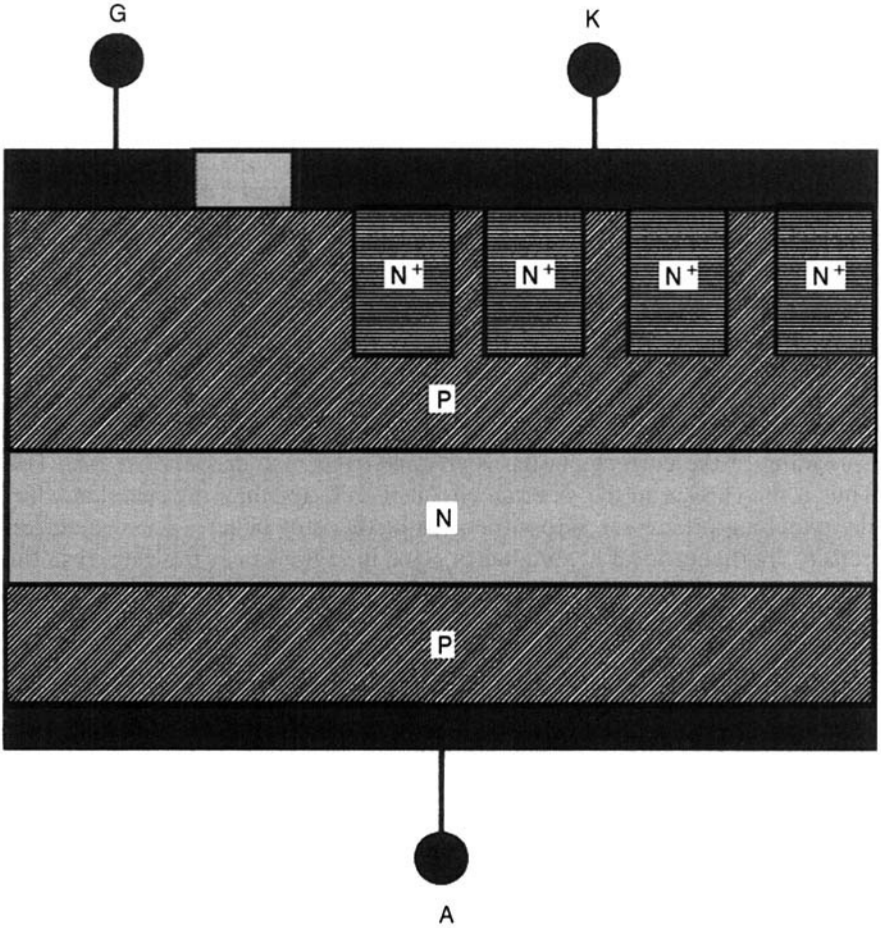


Figure 4.16 Shorted emitter construction of a thyristor to improve the *di/dt* capability.

Example 4.10 In an IGBT, the emitter of NPN transistor is short-circuited to its base by a resistance $R_b = 1\ \Omega$. If the current gain of the NPN transistor without short-circuiting ($R_b = \infty$) is α_{NPN} and its current gain with short-circuiting is $\alpha_{NPN,\,SC}$, show that $\alpha_{NPN,\,SC}$ is $\ll \alpha_{NPN}$ until the emitter–base junction of the NPN transistor is forward-biased by a voltage $V_{BE} > 0.7$ V. Given $\alpha_{NPN} = 0.99$ and reverse saturation current of emitter–base junction (I_0) is 10 pA.

Figure E4.10.1 gives the schematic representation of the IGBT for this example with the shorting resistor R_b across the emitter-base diode of N^+PN^- transistor. On forward-biasing the emitter–base junction of the NPN transistor at a voltage V_{BE}, the collector current is $I_C = I_0\ \{\exp(qV_{BE}/kT) - 1\}$. Also, the current flowing through the shorting resistor R_b is $I_R = V_{BE}/R_b$. The current flowing at the emitter terminal in the short-circuited transistor is $I_{E,\,SC} = I_C/\alpha_{NPN} + I_R$, where α_{NPN} is the current gain of the transistor without short-circuiting. So the current gain of the transistor

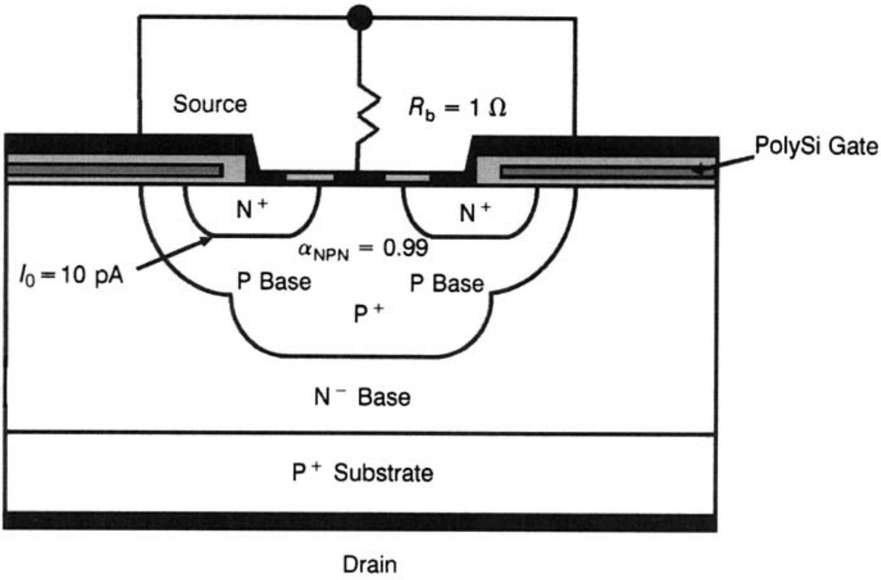


Figure E4.10.1 IGBT for Example 4.10.

with short-circuiting is $\alpha_{NPN,SC} = I_C/I_{E,SC} = I_C/(I_E/\alpha_{NPN} + I_R) = \alpha_{NPN}[1/\{1 + (I_R\alpha_{NPN})/I_E\}] = \alpha_{NPN}/[1 + (V_{BE}\ \alpha_{NPN})/\{R_b I_0 \exp(qV_{BE}/kT) - 1\}] = 0.99/[1 + (0.99\ V_{BE})/\{1 \times 10^{-11} \exp(qV_{BE}/kT) - 1\}]$. Putting $V_{BE} = 0.5$ V, we obtain $\alpha_{NPN,SC} = 0.99/[1 + (0.99 \times 0.5)/\{1 \times 10^{-11} \exp(0.5/0.0259) - 1\}] = 4.82 \times 10^{-3}$, which is $\ll \alpha_{NPN}$ $(= 0.99)$. Similarly for $V_{BE} = 0.6$ V, $\alpha_{NPN,SC} = 0.161$, which is again $< \alpha_{NPN}$ $(= 0.99)$. Now for $V_{BE} = 0.7$ V we obtain $\alpha_{NPN,SC} = 0.8786$, which is slightly less than α_{NPN} $(= 0.99)$. But for $V_{BE} = 0.8$ V, we obtain $\alpha_{NPN,SC} = 0.987$, which is close to α_{NPN} $(= 0.99)$. Thus starting from 0 V, as the emitter–base voltage increases, initially, most of the current flows through the resistor R_b and the current gain of the transistor is very small. But as soon as V_{BE} exceeds 0.7 V, the emitter–base junction is forward-biased, a large fraction of current flows through its emitter, and its current gain approaches that of the non-short-circuited transistor.

4.4.4 Thyristor Turn-on and Turn-off Times

Looking at Fig. 4.17a, if t_1 is the instant at which the gate voltage V_G rises to 0.1 of its final value and if t_2 and t_3 are the instants at which the thyristor current I rises to 0.1 and 0.9, respectively, of its final value, then we define the following: *Delay time* $t_d = t_2 - t_1$; rise time $t_r = t_3 - t_2$; and *turn-on time* $t_{ON} = t_d + t_r$.

The thyristor is turned off by applying a reverse voltage across its main terminals A and K. The turning off of a thyristor is similar to that of a diode, as explained in Section 4.1.5. So the turn-off waveforms of the thyristor (Fig.

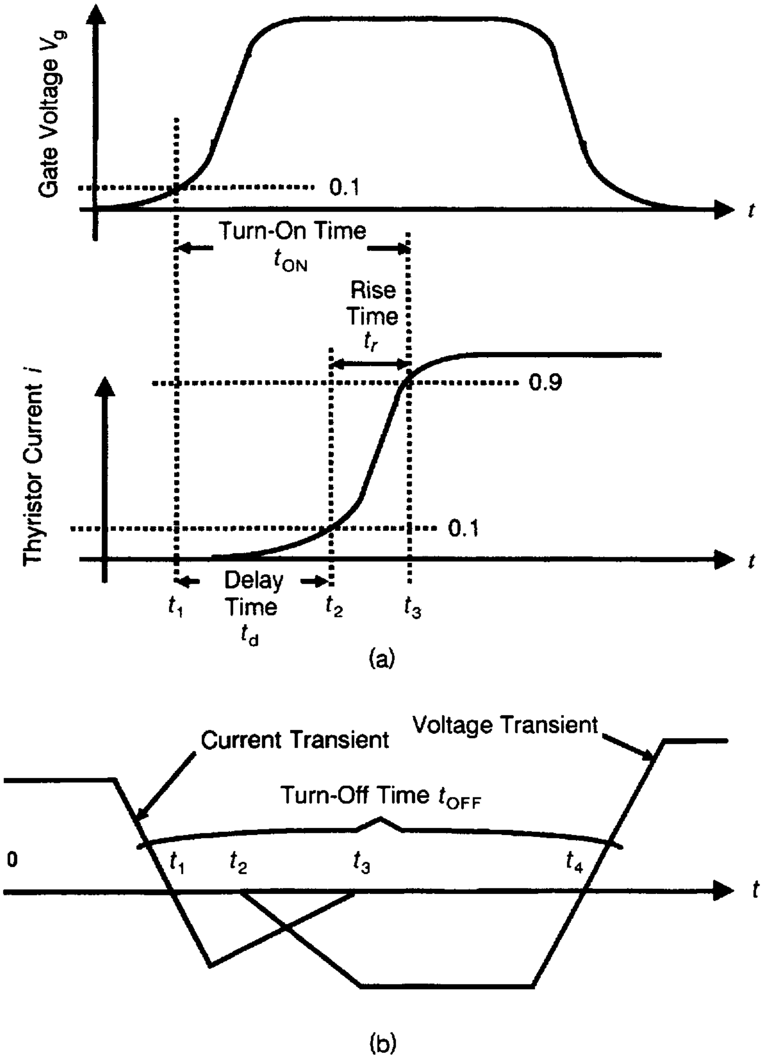


Figure 4.17 Thyristor switching waveforms: (a) Turn-on and (b) turn-off.

4.17b) resemble those of the diode (Fig. 4.5) up to the instant t_3. Turn-off switching of a thyristor is completed only after it has regained its ability to block forward voltage. But at instant t_3, the thyristor has only recovered its reverse blocking capability due to junction J_1. The forward blocking capability is due to junction J_2 and it takes a finite time for carriers to decay, by recombination, after the reverse recovery transient. Therefore, the forward blocking capability is recovered at the instant t_4 at which the voltage across the thyristor is zero and the device starts becoming forward-biased. The

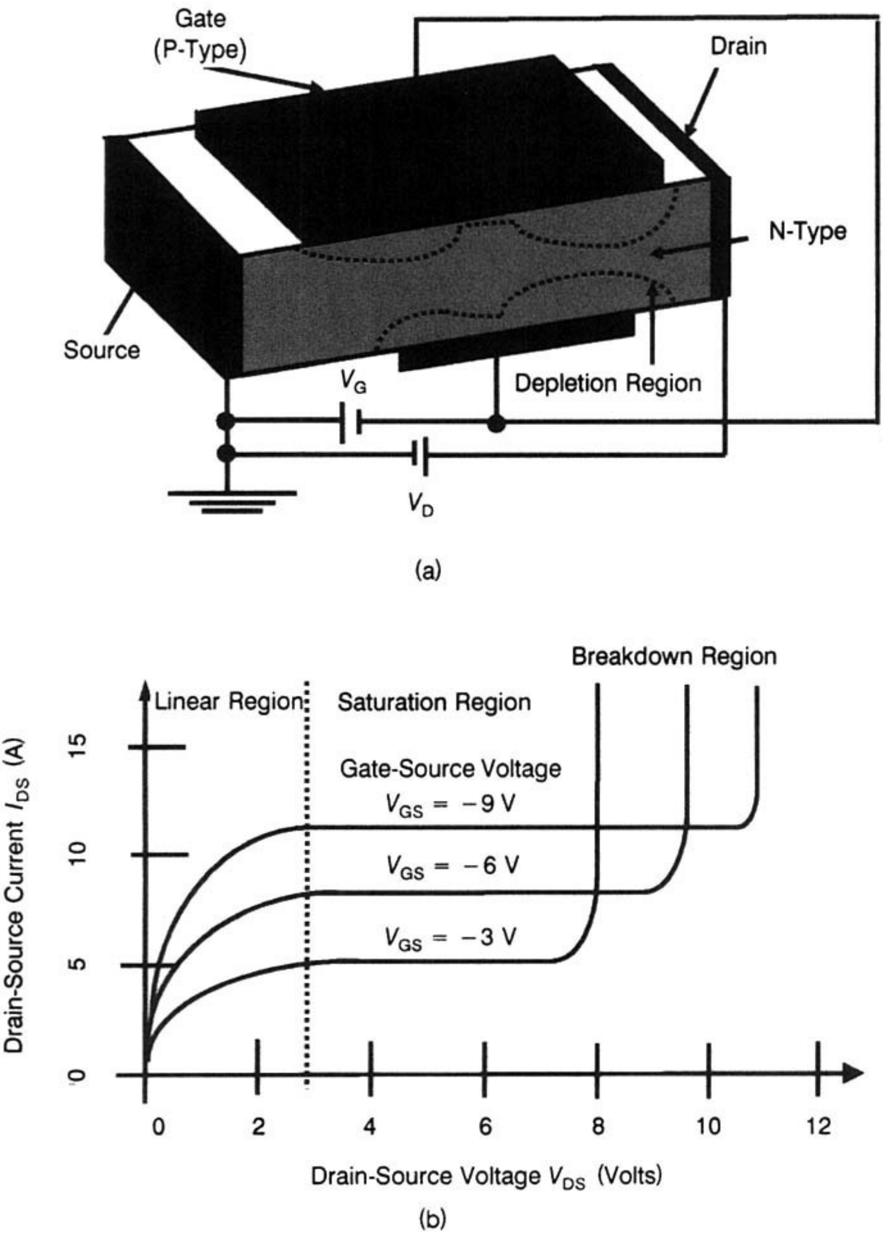

Gate
(P-Type)
Drain
N-Type
Source
Depletion Region
V_G
V_D
(a)
Breakdown Region
Linear Region
Saturation Region
Gate-Source Voltage
$V_{GS} = -9$ V
$V_{GS} = -6$ V
$V_{GS} = -3$ V
Drain-Source Current I_{DS} (A)
0
5
10
15
0
2
4
6
8
10
12
Drain-Source Voltage V_{DS} (Volts)
(b)

Figure 4.18 (a) JFET cross-sectional view. (b) Static characteristics of JFET showing the three operating regions.

turn-off time is $t_{\mathrm{OFF}} = t_4 - t_1$, where t_1 is the instant at which the forward current becomes zero. It has been demonstrated that application of a negative gate-to-cathode voltage speeds up the turn-off process of a thyristor.

4.5 JUNCTION FIELD-EFFECT TRANSISTOR (JFET)

The JFET (Fig. 4.18a) is a *voltage-controlled resistor* whose resistance is varied by changing the width of the depletion layer extending into the channel region [22–24]. It has three electrodes: the *source*, which acts as the origin of carriers; the *drain*, which is a carrier sink; and the *gate*, which controls the supply of carriers from source to drain. The source and drain make ohmic contacts with the channel while the gate forms a rectifying junction with it. The main device dimensions are: the *channel* or *gate length* (L), the *channel width* (W), *channel depth* a, *channel opening* b, and *depletion layer width* h.

Figure 4.18b shows the current–voltage characteristics of the JFET. These are a family of plots of drain–source current, i_{DS}, with respect to drain–source voltage v_{DS} for different values of the gate–source bias v_{GS}. The characteristics are divided into three regions: linear, saturation, and breakdown regions. In the *linear region*, $i_{\mathrm{DS}} \propto v_{\mathrm{DS}}$. In the *saturation region*, i_{DS} is constant and independent of v_{DS} whereas in the *breakdown region*, i_{DS} increases steeply with a small change of v_{DS}.

In IGBTs, the JFET effect appears as a component of the ON resistance, which was discussed in Section 3.4.1.

4.6 SUMMARIZING REMARKS

Embellishment of IGBT with various bipolar properties is fraught with both (a) their merits of lower static losses and higher breakdown voltages and (b) their demerits of larger dynamic losses. Every bipolar property must be understood in the background of its impact on IGBT behavior to appreciate how it can be molded to user's requirement.

REVIEW EXERCISES

4.1 Name the important specifications of a power diode.

4.2 What are the two main packages used for power diodes?

4.3 Derive an equation for the built-in potential of a P-N junction in terms of the dopant densities on the two sides.

4.4 Write the formulae for depletion layer width and capacitance of a reverse-biased P-N junction, explaining the symbols.

4.5 Give the equations for the avalanche breakdown voltage of abrupt and linearly graded junctions.

4.6 Derive the Shockley equation for the volt–ampere characteristic of a diode. Write the equation for the reverse leakage current showing its diffusion and generation components.

4.7 Draw the turn-off switching waveform of a P-N junction diode and explain the significance of each time segment in it. Define reverse recovery time of a diode.

4.8 Sketch the carrier distribution in a P-I-N rectifier. Indicate the regions of highest electron and hole concentrations and minimum carrier concentration.

4.9 Distinguish a double-diffused bipolar transistor from a triple-diffused type. What is the role of the N^+ layer on the backside of the latter.

4.10 How are the α and β of a transistor related? Write the formulae for the emitter injection efficiency γ and the base transport factor α_T of a transistor.

4.11 Explain the Webster's and Kirk's effects in a bipolar transistor. How does the current gain β of a transistor vary with the collector current? Give physical reasons for each portion of the β–I_C curve.

4.12 Express the minimum base current to maintain a saturated ON-state of the BJT in terms of its h_{FE}. What is "forced h_{FE}" operation of a transistor? What is the disadvantage of overdriving a transistor?

4.13 Draw a representative safe operating area (SOA) of a bipolar transistor and indicate the various boundaries on SOA. Point out the significance of each boundary. Explain the concept of sustaining voltage BV_{CES} of a transistor.

4.14 Draw the turn-on and turn-off switching transitions of a transistor. Mark the time instants for the different transitions taking place, and indicate the rise time, storage time, and fall times of the transistor.

4.15 How does a line-commutated thyristor differ from a force-commutated thyristor?

4.16 What are the three operating states of a thyristor? Show the biasing of the different internal junctions in these three states.

4.17 Representing a thyristor as a pair of PNP and NPN transistors, explain how regenerative action turns the device on.

4.18 Derive the equation describing the forward breakover of a thyristor when the sum of the current gains of its two transistors approaches unity.

4.19 How does the finite plasma spreading velocity impose a di/dt limitation on the thyristor. What special geometries are employed to improve the di/dt rating of a thyristor?

4.20 Explain the false turning on of a thyristor due to rapid application of voltage. How does a shorted emitter construction help in avoiding the thyristor turn-on due to dV/dt effect.

4.21 Sketch the thyristor turn-on and turn-off switching characteristics and explain the time delays experienced during both the transitions.

4.22 Calculate the built-in potential of the N^+-emitter/P-base junction of an IGBT in which the doping concentration of the N^+ emitter is 1.5×10^{20} cm^{-3} and that of the P base is 2.5×10^{17} cm^{-3}?

4.23 In an NPT-IGBT, the impurity doping concentration of the N^- base is 2×10^{14} cm^{-3} while that of the P base is 4×10^{17} cm^{-3}. The N^--base thickness is 180 μm. Find the (a) breakdown voltage, (b) the space-charge region thickness on N side, (c) the space-charge region thickness on P side, (d) the total space-charge region thickness, and (e) the undepleted length of the N^- base at 1000 V.

4.24 In an IGBT, the P^+-layer doping concentration is 1.9×10^{20} cm^{-3}, the N^--base concentration is 3×10^{14} cm^{-3}, and N^--base thickness is 75 μm. Electron diffusion constant in the emitter = 0.9 cm^2/sec, electron diffusion length in the emitter = 0.8 μm, hole diffusion constant in the base = 12 cm^2/sec, and hole lifetime in base = 7 μsec. Find the common-base current gain α_{PNP} of the P^+NP transistor in the IGBT.

REFERENCES

1. W. Shockley, The Theory of P-N Junctions in Semiconductors and P-N Junction Transistors, *Bell Sys. Tech. J.*, Vol. 28, No. 3, July 1949, pp. 435–489.
2. J. L. Moll, The Evolution of the Theory of the Current–Voltage Characteristics of P-N Junctions, *Proc. IRE*, Vol. 46, No. 6, June 1958, pp. 1076–1082.
3. R. N. Hall and W. C. Dunlap, P-N Junctions Prepared by Impurity Diffusion, *Phys. Rev.*, Vol. 80, 1950, p. 467.
4. S. K. Ghandhi, *Semiconductor Power Devices*, Wiley, New York, 1977.
5. S. M. Sze, *Physics of Semiconductor Devices*, Wiley Eastern Ltd., New Delhi, 1987.
6. J. Vithayathil, *Power Electronics: Principles and Applications*, McGraw-Hill, Inc., New York, 1995.
7. H. S. Veloric and M. B. Prince, High-Voltage Conductivity-Modulated Silicon Rectifier, *Bell Syst. Tech. J.*, Vol. 36, No. 4, July 1957, pp. 975–1004.
8. J. Bardeen and W. H. Brattain, The Transistor, A Semiconductor Triode, *Phys. Rev.*, Vol. 74, 1948, p. 230.
9. H. C. Theuerer, J. J. Kleimack, H. H. Loar, and H. Christenen, Epitaxial Diffused Transistors, *Proc. IRE*, Vol. 48, No. 9, September 1960, pp. 1642–1643.
10. B. J. Baliga and D. Y. Chen, *Power Transistors: Device Design and Applications*, IEEE Press, New York, 1984.
11. D. J. Roulston, *Bipolar Semiconductor Devices*, McGraw-Hill, New York, 1990.
12. A. Blicher, *Field Effect and Bipolar Power Transistor Physics*, Academic Press, New York, 1981.
13. W. M. Webster, On the Variation of Junction-Transistor Current Gain Amplification Factor with Emitter Current, *Proc. IRE*, Vol. 42, No. 6, June 1954, pp. 914–920.
14. C. T. Kirk, Jr., A Theory of Transistor Cut-off Frequency (f_T) Falloff at High Current Densities, *IRE Trans. Electron Devices*, Vol. ED-9, No. 2, March, 1962, pp. 164–173.
15. J. J. Ebers and J. L. Moll, Large Signal Behaviour of Junction Transistors, *Proc. IRE*, The Institute of Radio Engineers, Inc., New York, Vol. 42, No. 12, December 1954, pp. 1761–1772.
16. H. K. Gummel and H. C. Poon, An Integral Charge Control Model of Bipolar Transistors, *Bell Syst. Tech. J.*, Vol. 49, No. 5, May–June 1970, pp. 827–852.

17. A. Blicker, *Thyristor Physics*, Springer, New York, 1976.
18. A. Herlet, The Maximum Blocking Capability of Silicon Thyristors, *Solid State Electron.*, Vol. 8, 1965, pp. 655–671.
19. H. J. Ruhl, Jr., Spreading Velocity of the Active Area Boundary in a Thyristor, *IEEE Trans. Electron Devices*, Vol. ED-17, No. 9, September 1970, pp. 672–680.
20. H. G. Storm and J. G. St. Clair, An Involute Gate-Emitter Configuration for Thyristors, *IEEE Trans. Electron Devices*, Vol. ED-21, No. 8, August 1974, pp. 520–522.
21. P. S. Rederecht, A Review of the Shorted Emitter Principle as Applied to P-N-P-N Silicon Controlled Rectifiers, *Int. J. Electronics*, Vol. 31, No. 6, December 1971, pp. 541–564.
22. W. Shockley, A Unipolar Field-Effect Transistor, *Proc. IRE*, Vol. 40, No. 11, November 1952, pp. 1365–1376.
23. G. C. Dacey and I. M. Ross, Unipolar Field-Effect Transistor, *Proc. IRE*, Vol. 41, 1953, p. 970.
24. G. C. Dacey and I. M. Ross, The Field Effect Transistor, *Bell Syst. Tech. J.*, Vol. 34, No. 6, November 1955, pp. 1149–1189.

APPENDIX 4.1: DRIFT AND DIFFUSION CURRENT DENSITIES

In a semiconductor, carrier transport takes place by two main mechanisms, viz., drift under the influence of an applied electric field, and diffusion in a concentration gradient.

Carrier Drift

Suppose an electric field ξ is applied to a semiconductor bar of cross-sectional area A and length L having a carrier concentration of n electrons per unit volume (Fig. A4.1.1). If q is the charge and v_i is the velocity of electron, the electron current density is $-qv_i$. Adding together the contributions from the n electrons, the current density due to all the electrons contained in the

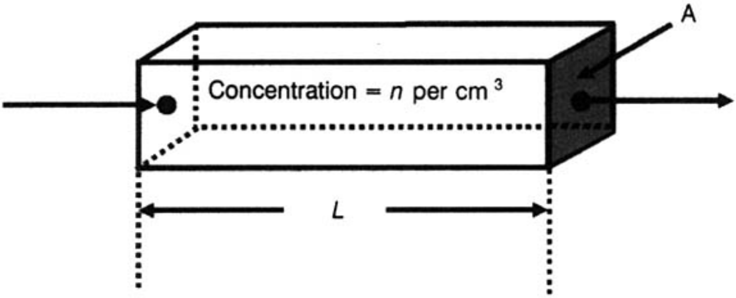


Figure A4.1.1 Electron drift in a uniformly doped N-type semiconductor bar.

unit volume is

$$J_n = \sum_{i=0}^{n} (-qv_i) = -qnv_n = qn\mu_n \xi \tag{A4.1.1}$$

where μ_n is the electron mobility (defined as the average drift velocity v_i per unit applied electric field ξ). A similar expression can be written for the holes remembering that the holes carry positive charge. Then, the current density due to holes is

$$J_p = qpv_p = qp\mu_p \xi \tag{A4.1.2}$$

where p is the number of holes per unit volume.

Combining together the electron and hole current densities given by Eqs. (A4.1.1) and (A4.1.2), the total current density J is

$$J = J_n + J_p = qn\mu_n \xi + qp\mu_p \xi = q\xi(n\mu_n + p\mu_p) \tag{A4.1.3}$$

where

$$n\mu_n + p\mu_p = \sigma$$

denotes the conductivity of the semiconductor, and its reciprocal

$$\rho = \frac{1}{n\mu_n + p\mu_p} \tag{A4.1.4}$$

is the resistivity.

Carrier Diffusion

This process takes place when there is a spatial variation of carrier concentration in the semiconductor so that the carriers move from the region of high concentration to one of low concentration. Let us consider a semiconductor specimen at a uniform temperature (Fig. A4.1.2). Suppose the specimen has a varying electron concentration $n(x)$ in the x-direction. The uniformity of temperature implies that the electrons have the same thermal energy. Let us calculate the average number of electrons crossing the plane at $x = 0$ per unit area per unit time. As we know, the electrons are in a state of random thermal motion with a thermal velocity v_{th} and mean free path $l = v_{th}\tau_c$, where τ_c is the mean free time. Since the electrons at the position $x = -l$, one mean free path away on the left, are equally likely to move toward the left or right, in time τ_c, half of these electrons will be able to cross the plane $x = 0$. Then the average rate of electrons crossing the plane $x = 0$ per unit area from the left side is

$$F_1 = \frac{1}{2}n(-l)\frac{l}{\tau_c} = \frac{1}{2}n(-l)\cdot v_{th} \tag{A4.1.5}$$

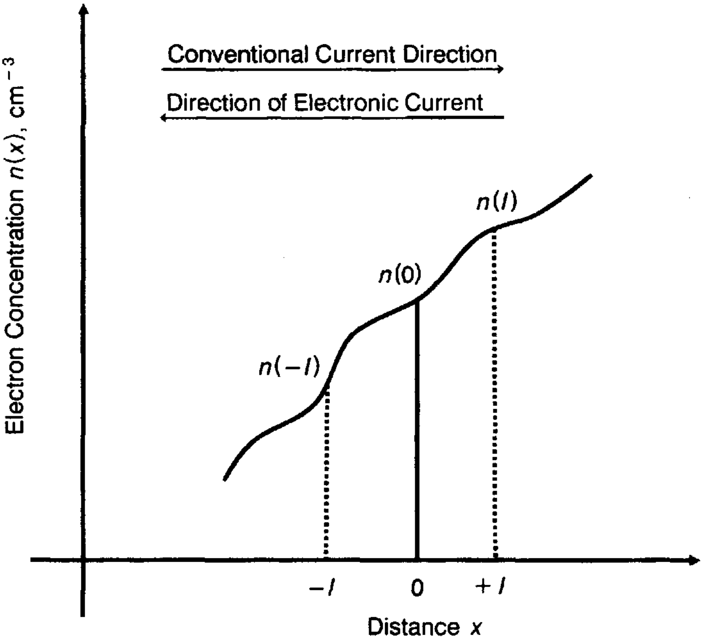

Conventional Current Direction
Direction of Electronic Current
Electron Concentration n(x), cm⁻³
n(l)
n(0)
n(−l)
−l
0
+l
Distance x

Figure A4.1.2 Variation of electron concentration with distance (l is the electron mean free path).

By Taylor series approximation, we have

$$n(-l) = n(0) - l\frac{dn}{dx} \tag{A4.1.6}$$

so that Eq. (A4.1.5) reduces to

$$F_1 = \frac{1}{2}\left\{n(0) - l\frac{dn}{dx}\right\}\cdot v_{\text{th}} \tag{A4.1.7}$$

Similarly, the rate of electron flow crossing the plane $x = l$ from the right side is

$$F_2 = \frac{1}{2}n(l)\frac{l}{\tau_c} = \frac{1}{2}n(l)\cdot v_{\text{th}} \tag{A4.1.8}$$

which becomes

$$F_2 = \frac{1}{2}\left\{n(0) + l\frac{dn}{dx}\right\}\cdot v_{\text{th}} \tag{A4.1.9}$$

on applying the Taylor series. From Eqs. (A4.1.7) and (A4.1.9), the net rate of electron flow from left to right is

$$F_1 - F_2 = F = \frac{1}{2}\left\{n(0) - l\frac{dn}{dx}\right\}\cdot v_{\mathrm{th}} - \frac{1}{2}\left\{n(0) + l\frac{dn}{dx}\right\}\cdot v_{\mathrm{th}} = -v_{\mathrm{th}} l\frac{dn}{dx} \tag{A4.1.10}$$

This equation is recast as

$$F = -v_{\mathrm{th}} l\frac{dn}{dx} = -D_{\mathrm{n}}\frac{dn}{dx} \tag{A4.1.11}$$

where $D_{\mathrm{n}} = v_{\mathrm{th}} l$ is a constant known as the *diffusion constant* or *diffusivity*. Because of the charge $-q$ carried by an electron, the above electron flow results in an electron current density

$$J_{\mathrm{n}} = -qF = qD_{\mathrm{n}}\frac{dn}{dx} \tag{A4.1.12}$$

Proceeding in identical fashion, the spatial variation of hole concentration produces a corresponding hole current density

$$J_{\mathrm{p}} = -qD_{\mathrm{p}}\frac{dp}{dx} \tag{A4.1.13}$$

Total Current Density Due to Drift and Diffusion

Applying Eqs. (A4.1.1) and (A4.1.12), for electrons, we obtain

$$J_{\mathrm{n}} = qn\mu_{\mathrm{n}}\,\xi + qD_{\mathrm{n}}\frac{dn}{dx} \tag{A4.1.14}$$

Similarly, from Eqs. (A4.1.2) and (A4.1.14), we get for holes

$$J_{\mathrm{p}} = qn\mu_{\mathrm{p}}\,\xi - qD_{\mathrm{p}}\frac{dp}{dx} \tag{A4.1.15}$$

APPENDIX 4.2: EINSTEIN'S EQUATION

It is an important equation interrelating the two major phenomena of carrier transport in semiconductors, drift and diffusion, characterized by the parameters *mobility* and *diffusion constant*, respectively.

From the law of equipartition of energy, we can write for the one-dimensional case

$$\tfrac{1}{2}m_{\mathrm{n}}{v_{\mathrm{th}}}^2 = \tfrac{1}{2}kT \tag{A4.2.1}$$

where m_n is the mass of electron, k is Boltzmann's constant, and T is the absolute temperature.

To calculate the drift velocity v_d of an electron in an applied electric field ξ, the impulse (product of force and time) applied to the electron during its free flight between successive collisions (time = τ) is equated to the momentum gained by the electron during the above period. Since the force acting on the electron equals $-q\xi$ and the momentum acquired by it is $m_n v_d$, we have

$$-q\xi\tau = m_n v_d \tag{A4.2.2}$$

which gives

$$v_d = \frac{-q\xi\tau}{m_n} \tag{A4.2.3}$$

From this equation, the drift velocity per unit applied electric field $(-v_d/\xi)$ or the electron mobility μ_n is

$$\mu_n = -\frac{v_d}{\xi} = \frac{q\tau}{m_n} \tag{A4.2.4}$$

Substituting for m_n from Eq. (A4.2.1) into Eq. (A4.2.4), we get

$$\mu_n = -\frac{v_d}{\xi} = \frac{q\tau}{kT/v_{th}^2} = \frac{v_{th}^2 q\tau}{kT} \tag{A4.2.5}$$

But $\tau = l/v_{th}$, so Eq. (A4.2.5) reduces to

$$\mu_n = -\frac{v_d}{\xi} = \frac{q\tau}{kTv_{th}^2} = \frac{v_{th}^2 ql}{kTv_{th}} = \frac{v_{th} ql}{kT} \tag{A4.2.6}$$

Substituting $v_{th} l = D_n$ from Eq. (A4.1.11) into Eq. (A4.2.6), we have

$$\mu_n = \frac{D_n q}{kT} \quad \text{or} \quad \frac{D_n}{\mu_n} = \frac{kT}{q} \tag{A4.2.7}$$

A similar equation applies to holes:

$$\frac{D_p}{\mu_p} = \frac{kT}{q} \tag{A4.2.8}$$

so that we can write, in general,

$$\frac{D}{\mu} = \frac{kT}{q} \tag{A4.2.9}$$

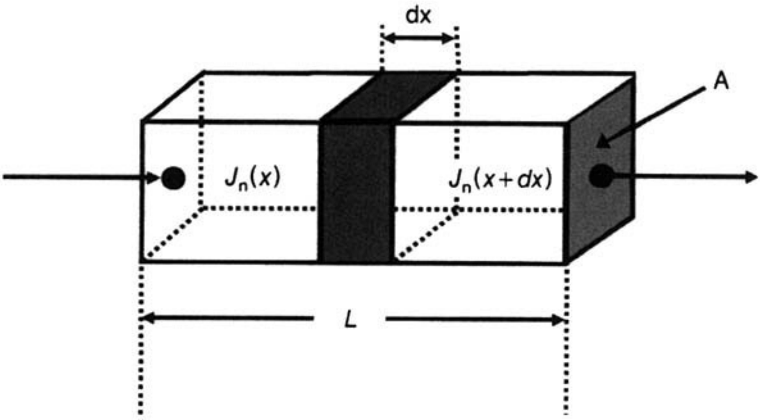


Figure A4.3.1 Infinitesimal slice of thickness *dx*.

APPENDIX 4.3: CONTINUITY EQUATION AND ITS SOLUTION

It is the overall governing equation for carrier drift, diffusion, and generation–recombination mechanisms in a semiconductor. Considering an infinitesimal slice (Fig. A4.3.1) of cross-sectional area A and thickness dx located at x, the rate of change in the number electrons within the slice is expressed as (Number of electrons entering the slice per unit time at x) − (Number of electrons leaving the slice per unit time at $x + dx$) + (Rate of electron generation in the slice − Rate of electron recombination in the slice). This equality is mathematically expressed as

$$\frac{\partial n}{\partial t} A\,dx = \left\{ \frac{J_n(x)A}{-q} - \frac{J_n(x+dx)A}{-q} \right\} + (G_n - R_n)A\,dx \quad \text{(A4.3.1)}$$

where G_n is the generation rate and R_n is the recombination rate. By Taylor series expansion,

$$J_n(x+dx) = J_n(x) + \frac{\partial J_n}{\partial x} dx + \cdots \quad \text{(A4.3.2)}$$

From Eqs. (A4.3.1) and (A4.3.2), we have

$$\frac{\partial n}{\partial t} = \frac{1}{q} \frac{\partial J_n}{\partial x} + (G_n - R_n) \quad \text{(A4.3.3)}$$

which is the continuity equation for electrons. Proceeding in the same manner, the continuity equation for holes is

$$\frac{\partial p}{\partial t} = -\frac{1}{q} \frac{\partial J_p}{\partial x} + (G_p - R_p) \quad \text{(A4.3.4)}$$

Now, in a semiconductor, the recombination rate of excess minority carriers is proportional to the concentration of excess carriers. The constant of proportionality is called the *excess carrier lifetime* (τ). The recombination rate of minority-carrier electrons (R_n) varies as the excess electron concentration ($n_p - n_{p0}$) where n_p is the instantaneous electron concentration and n_{p0} is the electron concentration at thermal equilibrium. It is expressed as

$$R_n = \frac{n_p - n_{p0}}{\tau_n} \tag{A4.3.5}$$

where the constant of proportionality is the *electron lifetime* τ_n. The equation for the hole recombination rate R_p is

$$R_p = \frac{p_n - p_{n0}}{\tau_p} \tag{A4.3.6}$$

where τ_p is the hole lifetime. Applying eqs. (A4.3.5) and (A4.3.6) and the equations for electron and hole current densities J_n and J_p, given in Appendix 4.1, Eqs. (A4.3.3) and (A4.3.4) are modified as

$$\frac{\partial n_p}{\partial t} = \mu_n n_p \frac{\partial \xi}{\partial x} + \mu_n \xi \frac{\partial n_p}{\partial x} + D_n \frac{\partial^2 n_p}{\partial x^2} + G_n - \frac{n_p - n_{p0}}{\tau_n} \tag{A4.3.7}$$

$$\frac{\partial p_n}{\partial t} = -\mu_p p_n \frac{\partial \xi}{\partial x} - \mu_p \xi \frac{\partial p_n}{\partial x} + D_p \frac{\partial^2 p_n}{\partial x^2} + G_p - \frac{p_n - p_{n0}}{\tau_p} \tag{A4.3.8}$$

Equations (A4.3.7) and (A4.3.8) are the continuity equations for electrons and holes, respectively.

Let us focus our attention on Eq. (A4.3.8) under the assumption that no generation current exists in the depletion region, that is, $G_p = 0$. Furthermore, in the neutral N or P regions the electric field $\xi = 0$. For the steady state, $\partial p_n / \partial t = 0$. Then Eq. (A4.3.8) is simplified to

$$D_p \frac{\partial^2 p_n}{\partial x^2} - \frac{p_n - p_{n0}}{\tau_p} = 0 \tag{A4.3.9}$$

or

$$\frac{\partial^2 p_n}{\partial x^2} - \frac{p_n - p_{n0}}{D_p \tau_p} = 0 \quad \therefore \frac{\partial^2 p_n}{\partial x^2} - \frac{p_n - p_{n0}}{L_p^2} = 0 \tag{A4.3.10}$$

where $L_p = \sqrt{(D_p \tau_p)}$ is the diffusion length of holes.

To solve this equation, we note that it is of the standard form

$$aD^2 y + bDy + cy = f(x) \tag{A4.3.11}$$

where a, b, and c are constants, $D = d/dx$, $D^2 = d^2/dx^2$. Its characteristic equation is

$$m^2 - \frac{1}{L_p^2} = 0 \tag{A4.3.12}$$

Hence $m = \pm 1/L_p$, and the solution is

$$p_n = A\exp(m_1 x) + B\exp(m_2 x) \tag{A4.3.13}$$

where $m_1 = +1/L_p$ and $m_2 = -1/L_p$ are the two values of m. Putting the values of m_1 and m_2 into Eq. (A4.3.13), we get

$$p_n = A\exp\left(\frac{x}{L_p}\right) + B\exp\left(-\frac{x}{L_p}\right) \tag{A4.3.14}$$

Because p_n cannot increase with x, the first term is physically untenable so the solution of Eq. (A4.3.10) reduces to

$$p_n = B\exp\left(-\frac{x}{L_p}\right) \tag{A4.3.14}$$

With the help of the boundary condition given by Eq. (4.24), at $x = 0$ we have

$$p_n = p_n(0) = p_{n0}\exp\left(\frac{v_a}{\phi}\right) = B\exp(0) = B \tag{A4.3.15}$$

Substituting the value of B in Eq. (A4.3.14), we have

$$p_n = p_n(0)\exp\left(-\frac{x}{L_p}\right) \tag{A4.3.16}$$

Hence, for the decay of minority carriers away from the junction we may write

$$p_n'(x) = p_n'(0)\exp\left(-\frac{x}{L_p}\right) \tag{A4.3.17}$$

where, as defined by Eq. (4.27), we obtain

$$p_n'(x) = p_n(x) - p_{n0} \tag{A4.3.18}$$

Applying Eq. (A4.3.18) to $x = 0$,

$$p_n'(0) = p_n(0) - p_{n0} \tag{A4.3.19}$$

Substituting for $p_n'(0)$ from Eq. (A4.3.19) into Eq. (A4.3.17),

$$p_n'(x) = p_n'(0)\exp\left(-\frac{x}{L_p}\right) = \{p_n(0) - p_{n0}\}\exp\left(-\frac{x}{L_p}\right) \tag{A4.3.20}$$

Now putting the value of $p_n(0)$ from Eq. (A4.3.15) into Eq. (A4.3.20), the result is

$$\begin{aligned} p_n'(x) &= \left\{p_{n0}\exp\left(\frac{v_a}{\phi}\right) - p_{n0}\right\}\exp\left(-\frac{x}{L_p}\right) \\ &= p_{n0}\left\{\exp\left(\frac{v_a}{\phi}\right) - 1\right\}\exp\left(-\frac{x}{L_p}\right) \end{aligned} \tag{A4.3.21}$$

which is Eq. (4.30).

APPENDIX 4.4: SOLUTION OF THE CONTINUITY EQUATION (4.41)

The general solution to the differential equation (4.41) is

$$n(x) = A\cosh\left(\frac{x}{L_a}\right) + B\sinh\left(\frac{x}{L_a}\right) \tag{A4.4.1}$$

where A and B are constants determined from the boundary conditions, Eqs. (4.44) and (4.45), which may be rewritten as

$$\left.\frac{dn}{dx}\right|_{x=+d} = \frac{J}{2qD_n} \tag{A4.4.2}$$

$$\left.\frac{dn}{dx}\right|_{x=-d} = -\frac{J}{2qD_p} \tag{A4.4.3}$$

By differentiation of Eqs. (A4.4.1) at $x = +d$ and $-d$, the following equations are obtained:

$$\left.\frac{dn}{dx}\right|_{x=+d} = \frac{1}{L_a}\left\{A\sinh\left(\frac{d}{L_a}\right) + B\cosh\left(\frac{d}{L_a}\right)\right\} \tag{A4.4.4}$$

$$\left.\frac{dn}{dx}\right|_{x=-d} = \frac{1}{L_a}\left\{-A\sinh\left(\frac{d}{L_a}\right) + B\cosh\left(\frac{d}{L_a}\right)\right\} \tag{A4.4.5}$$

Equating the right-hand side of Eq. (A4.4.4) with Eq. (A4.4.2) and that of Eq. (A4.4.5) with Eq. (A4.4.3), we get

$$A\sinh\left(\frac{d}{L_a}\right)+B\cosh\left(\frac{d}{L_a}\right)=\frac{JL_a}{2qD_n} \tag{A4.4.6}$$

$$-A\sinh\left(\frac{d}{L_a}\right)+B\cosh\left(\frac{d}{L_a}\right)=-\frac{JL_a}{2qD_p} \tag{A4.4.7}$$

Adding together Eqs. (A4.4.6) and (A4.4.7), we obtain

$$2B\cosh\left(\frac{d}{L_a}\right)=\frac{JL_a}{2q}\left(\frac{D_p-D_n}{D_nD_p}\right) \tag{A4.4.8}$$

from which

$$B=-\frac{JL_a(D_n-D_p)}{4qD_nD_p\cosh(d/L_a)} \tag{A4.4.9}$$

Subtracting Eq. (A4.4.7) from Eq. (A4.4.6), we get

$$2A\sinh\left(\frac{d}{L_a}\right)=\frac{JL_a}{2q}\left(\frac{D_n+D_p}{D_nD_p}\right) \tag{A4.4.10}$$

giving

$$A=\frac{JL_a(D_p+D_n)}{4qD_nD_p\sinh(d/L_a)} \tag{A4.4.11}$$

Substituting the values of the constants A and B from Eqs. (A4.4.11) and (A4.4.9) into Eq. (A4.4.1), we obtain

$$\begin{aligned}n(x)&=\frac{JL_a(D_n+D_p)}{4qD_nD_p\sinh(d/L_a)}\cosh\left(\frac{x}{L_a}\right)-\frac{JL_a(D_n-D_p)}{4qD_nD_p\cosh(d/L_a)}\sinh\left(\frac{x}{L_a}\right)\\&=\frac{JL_a}{4q}\left(\frac{D_n+D_p}{D_nD_p}\right)\left\{\frac{\cosh\left(\frac{x}{L_a}\right)}{\sinh\left(\frac{d}{L_a}\right)}-\left(\frac{D_n-D_p}{D_n+D_p}\right)\frac{\sinh\left(\frac{x}{L_a}\right)}{\cosh\left(\frac{d}{L_a}\right)}\right\}\end{aligned} \tag{A4.4.12}$$

In this equation,

$$\frac{JL_\mathrm{a}}{4q}\left(\frac{D_\mathrm{n}+D_\mathrm{p}}{D_\mathrm{n}D_\mathrm{p}}\right)=\frac{JL_\mathrm{a}{}^2}{4qL_\mathrm{a}}\left(\frac{D_\mathrm{n}+D_\mathrm{p}}{D_\mathrm{n}D_\mathrm{p}}\right)=\frac{JD_\mathrm{a}\tau_\mathrm{a}}{4qL_\mathrm{a}}\left(\frac{D_\mathrm{n}+D_\mathrm{p}}{D_\mathrm{n}D_\mathrm{p}}\right) \tag{A4.4.13}$$

using the relation

$$L_\mathrm{a}=\sqrt{D_\mathrm{a}\tau_\mathrm{a}} \tag{A4.4.14}$$

where τ_a is the ambipolar lifetime and D_a is the ambipolar diffusion coefficient given by

$$D_\mathrm{a}=\frac{n+p}{n/D_\mathrm{p}+p/D_\mathrm{n}}=\frac{2n}{n/D_\mathrm{p}+n/D_\mathrm{n}}=\frac{2D_\mathrm{n}D_\mathrm{p}}{D_\mathrm{n}+D_\mathrm{p}} \tag{A4.4.15}$$

since $n=p$. Also,

$$\frac{D_\mathrm{n}-D_\mathrm{p}}{D_\mathrm{n}+D_\mathrm{p}}\cong\frac{3D_\mathrm{p}-D_\mathrm{p}}{3D_\mathrm{p}+D_\mathrm{p}}=\frac{2D_\mathrm{p}}{4D_\mathrm{p}}=\frac{1}{2} \tag{A4.4.16}$$

because the electron diffusion coefficient is approximately three times the hole diffusion coefficient. Applying Eqs. (A4.4.15) and (A4.4.16) to Eq. (A4.4.13), we get

$$\frac{JD_\mathrm{a}\tau_\mathrm{a}}{2qL_\mathrm{a}}\left(\frac{D_\mathrm{n}+D_\mathrm{p}}{D_\mathrm{n}D_\mathrm{p}}\right)=\frac{J\left(\dfrac{2D_\mathrm{n}D_\mathrm{p}}{D_\mathrm{n}+D_\mathrm{p}}\right)\tau_\mathrm{a}}{4qL_\mathrm{a}}\left(\frac{D_\mathrm{n}+D_\mathrm{p}}{D_\mathrm{n}D_\mathrm{p}}\right)=\frac{J\tau_\mathrm{a}}{2qL_\mathrm{a}} \tag{A4.4.17}$$

From Eqs. (A4.4.12), (A4.4.13), (A4.4.16) and (A4.4.17), we obtain

$$n(x)=\frac{J\tau_\mathrm{a}}{2qL_\mathrm{a}}\left\{\frac{\cosh\left(\dfrac{x}{L_\mathrm{a}}\right)}{\sinh\left(\dfrac{d}{L_\mathrm{a}}\right)}-\frac{1}{2}\frac{\sinh\left(\dfrac{x}{L_\mathrm{a}}\right)}{\cosh\left(\dfrac{d}{L_\mathrm{a}}\right)}\right\} \tag{A4.4.18}$$

APPENDIX 4.5: DERIVATION OF EQ. (4.50)

From Eqs. (4.46) and (4.49) we have

$$V_{\mathrm{N}-} = \int_{-d}^{+d} \xi \, dx = \int_{-d}^{+d} \frac{J\,dx}{q(\mu_{\mathrm{n}}+\mu_{\mathrm{p}})n} - \frac{kT}{2q}\int_{-d}^{+d}\frac{dn}{n} = \frac{2qL_{\mathrm{a}}}{\tau_{\mathrm{HL}}J}\,\frac{J}{q(\mu_{\mathrm{n}}+\mu_{\mathrm{p}})}$$

$$\times \int_{-d}^{+d} \frac{dx}{\left\{\dfrac{\cosh(x/L_{\mathrm{a}})}{\sinh(d/L_{\mathrm{a}})} - \dfrac{1}{2}\,\dfrac{\sinh(x/L_{\mathrm{a}})}{\cosh(d/L_{\mathrm{a}})}\right\}} - \frac{kT}{2q}\int_{-d}^{+d}\frac{dn}{n} \qquad \text{(A4.5.1)}$$

Let us first confine our attention on the first term of this equation. The pre-integral expression in the first term is

$$= \frac{2L_{\mathrm{a}}^{2}}{\tau_{\mathrm{HL}}L_{\mathrm{a}}}\,\frac{1}{(\mu_{\mathrm{n}}+\mu_{\mathrm{p}})} = \frac{2D_{\mathrm{a}}\tau_{\mathrm{a}}}{\tau_{\mathrm{HL}}L_{\mathrm{a}}}\,\frac{1}{(\mu_{\mathrm{n}}+\mu_{\mathrm{p}})} = \frac{2D_{\mathrm{a}}}{L_{\mathrm{a}}(\mu_{\mathrm{n}}+\mu_{\mathrm{p}})} \qquad \text{(A4.5.2)}$$

where the substitutions $L_{\mathrm{a}} = \sqrt{(D_{\mathrm{a}}\tau_{\mathrm{a}})}$ and $\tau_{\mathrm{HL}} = \tau_{\mathrm{a}}$ have been used. Combining Eqs. (A4.5.1) and (A4.5.2) we have

$$V_{\mathrm{N}-} = \frac{2D_{\mathrm{a}}}{L_{\mathrm{a}}(\mu_{\mathrm{n}}+\mu_{\mathrm{p}})}\int_{-d}^{+d} \frac{dx}{\left\{\dfrac{\cosh(x/L_{\mathrm{a}})}{\sinh(d/L_{\mathrm{a}})} - \dfrac{1}{2}\,\dfrac{\sinh(x/L_{\mathrm{a}})}{\cosh(d/L_{\mathrm{a}})}\right\}} - \frac{kT}{2q}\int_{-d}^{+d}\frac{dn}{n} \qquad \text{(A4.5.3)}$$

The first term of Eq. (A4.5.3) is expressed as

$$V_{\mathrm{N}-} = \frac{2D_{\mathrm{a}}}{L_{\mathrm{a}}(\mu_{\mathrm{a}}+\mu_{\mathrm{p}})} \times$$

$$\int_{-d}^{+d} \frac{dx}{\left\{\dfrac{\cosh(x/L_{\mathrm{a}})\cosh(d/L_{\mathrm{a}}) - \frac{1}{2}\sinh(x/L_{\mathrm{a}})\sinh(d/L_{\mathrm{a}})}{\sinh(d/L_{\mathrm{a}})\cosh(d/L_{\mathrm{a}})}\right\}}$$

$$= \frac{2D_{\mathrm{a}}}{L_{\mathrm{a}}(\mu_{\mathrm{n}}+\mu_{\mathrm{p}})} \times$$

$$\int_{-d}^{+d} \frac{\sinh(d/L_{\mathrm{a}})\cosh(d/L_{\mathrm{a}})\,dx}{\left\{\cosh(x/L_{\mathrm{a}})\cosh(d/L_{\mathrm{a}}) - \frac{1}{2}\sinh(x/L_{\mathrm{a}})\sinh(d/L_{\mathrm{a}})\right\}}$$

$$= \frac{2D_{\mathrm{a}}\sinh(d/L_{\mathrm{a}})}{L_{\mathrm{a}}(\mu_{\mathrm{n}}+\mu_{\mathrm{p}})}\int_{-d}^{+d} \frac{dx}{\left\{\cosh(x/L_{\mathrm{a}}) - \frac{1}{2}\sinh(x/L_{\mathrm{a}})\tanh(d/L_{\mathrm{a}})\right\}} \qquad \text{(A4.5.4)}$$

This integral is converted into the form $\int dz/(z^2+a^2) = (1/a)\tan^{-1}(z/a) +$ integration constant, by a transformation of variables achieved by putting $\tanh(x/2L_a) = t$. Then $\operatorname{cosech}^2\{x/(2L_a)\} = \coth^2\{x/(2L_a)\} - 1 = 1/\tanh^2\{x/(2L_a)\} - 1 = 1/t^2 - 1 = (1-t^2)/t^2$. Therefore, $\operatorname{cosech}^2\{x/(2L_a)\} = \surd(1-t^2)/t$, and $\sinh\{x/(2L_a)\} = t/\surd(1-t^2)$. Now, $\cosh^2\{x/(2L_a)\} - \sinh^2\{x/(2L_a)\} = 1$. $\therefore \cosh^2\{x/(2L_a)\} = 1 + \sinh^2\{x/(2L_a)\} = 1 + t^2/(1-t^2) = (1 - t^2 + t^2)/(1-t^2) = 1/(1-t^2)$. So, $\cosh\{x/(2L_a)\} = 1/\surd(1-t^2)$, giving $\sinh(x/L_a) = 2\sinh\{x/(2L_a)\}\cosh\{x/(2L_a)\} = 2\{t/\surd(1-t^2)\}\{1/\surd(1-t^2)\} = 2t/(1-t^2)$. Also, $(\cosh x/L_a) = \cosh^2\{x/(2L_a)\} + \sinh^2\{x/(2L_a)\} = 1/(1-t^2) + t^2/(1-t^2) = (1+t^2)/(1-t^2)$. Again, $dt/dx = (d/dx)[\tanh\{x/(2L_a)\}] = (1/2L_a)[\operatorname{sech}^2\{x/(2L_a)\}] = (1/2L_a)[[1 - \tanh^2\{x/(2L_a)\}] = (1/2L_a)(1-t^2)$. $\therefore dx/dt = 2/(1-t^2)$. $\therefore dx = 2L_a dt/(1-t^2)$. Substituting the values of $\cosh(x/L_a)$, $\sinh(x/L_a)$ and dx in Eq. (A4.2.4), we find that the first term of Eq. (A4.5.3) is

$$= \frac{4D_a\sinh(d/L_a)}{(\mu_n+\mu_p)}\int_{-d}^{+d}\frac{\dfrac{dt}{1-t^2}}{\left\{\dfrac{1+t^2}{1-t^2} - \dfrac{1}{2}\tanh(d/L_a)\times\dfrac{2t}{1-t^2}\right\}}$$

$$= \frac{4D_a\sinh(d/L_a)}{(\mu_n+\mu_p)}\int_{-d}^{+d}\frac{dt}{1+t^2-t\times\tanh(d/L_a)} \qquad \text{(A4.5.5)}$$

The integral portion of Eq. (A4.5.5) is altered into the form

$$\int_{-d}^{+d}\frac{dt}{1+t^2-t\times\tanh(d/L_a)}$$

$$= \int_{-d}^{+d}\frac{dt}{\left[t^2 - 2t\times\frac{1}{2}\tanh(d/L_a) + \left\{\frac{1}{2}\tanh(d/L_a)\right\}^2 + 1 - \left\{\frac{1}{2}\tanh(d/L_a)\right\}^2\right]}$$

$$= \int_{-d}^{+d}\frac{dt}{\left[\left\{t - \frac{1}{2}\tanh(d/L_a)\right\}^2 + \left\{\sqrt{1-\frac{1}{4}\tanh^2(d/L_a)}\right\}^2\right]}$$

$$= \frac{1}{\sqrt{1-\frac{1}{4}\tanh^2(d/L_a)}}\left[\tan^{-1}\left\{\frac{t-\frac{1}{2}\tanh(d/L_a)}{\sqrt{1-\frac{1}{4}\tanh^2(d/L_a)}}\right\}\right]_{-d}^{+d} \qquad \text{(A4.5.6)}$$

Equation (A4.5.5) becomes

$$= \frac{4D_a \sinh(d/L_a)}{(\mu_n + \mu_p)} \frac{1}{\sqrt{1 - \frac{1}{4}\tanh^2(d/L_a)}} \left[\tan^{-1}\left\{\frac{t - \frac{1}{2}\tanh(d/L_a)}{\sqrt{1 - \frac{1}{4}\tanh^2(d/L_a)}}\right\}\right]_{-d}^{+d}$$

$$= \frac{4D_a \sinh(d/L_a)}{(\mu_n + \mu_p)} \frac{1}{\sqrt{1 - \frac{1}{4}\tanh^2(d/L_a)}}$$

$$\times \left[\tan^{-1}\left\{\frac{\tanh\left(\dfrac{x}{2L_a}\right) - \frac{1}{2}\tanh(d/L_a)}{\sqrt{1 - \frac{1}{4}\tanh^2(d/L_a)}}\right\}\right]_{-d}^{+d}$$

$$= \frac{4D_a \sinh(d/L_a)}{(\mu_n + \mu_p)} \frac{1}{\sqrt{1 - \frac{1}{4}\tanh^2(d/L_a)}} \times$$

$$\tan^{-1}\left[\frac{\dfrac{\tanh\left(\dfrac{d}{2L_a}\right) - \frac{1}{2}\tanh(d/L_a) - \left\{-\tanh\left(\dfrac{d}{2L_a}\right) - \frac{1}{2}\tanh(d/L_a)\right\}}{\sqrt{1 - \frac{1}{4}\tanh^2(d/L_a)}}}{1 - \dfrac{\tanh^2\{d/(2L_a)\} - \frac{1}{4}\tanh^2(d/L_a)}{1 - \frac{1}{4}\tanh^2(d/L_a)}}\right] \tag{A4.5.7}$$

by applying the formula $\tan^{-1} A - \tan^{-1} B = \tan^{-1}\{(A - B)/(1 - AB)$. On further simplification, Eq. (A4.5.7) reduces to

$$= \frac{4D_a \sinh(d/L_a)}{(\mu_n + \mu_p)} \frac{1}{\sqrt{1 - \frac{1}{4}\tanh^2(d/L_a)}} \times$$

$$\tan^{-1}\left[\frac{\dfrac{2\tanh\left(\dfrac{d}{2L_a}\right)}{\sqrt{1 - \frac{1}{4}\tanh^2(d/L_a)}}}{\dfrac{1 - \dfrac{1}{4}\tanh^2(d/L_a) - \tanh^2\{d/(2L_a)\} + \frac{1}{4}\tanh(d/L_a)}{1 - \frac{1}{4}\tanh^2(d/L_a)}}\right]$$

$$= \frac{4D_\mathrm{a}\sinh(d/L_\mathrm{a})}{(\mu_\mathrm{n}+\mu_\mathrm{p})}\frac{1}{\sqrt{1-\frac{1}{4}\tanh^2(d/L_\mathrm{a})}}\times$$

$$\tan^{-1}\left[\frac{2\tanh\left(\dfrac{d}{2L_\mathrm{a}}\right)\sqrt{1-\frac{1}{4}\tanh^2(d/L_\mathrm{a})}}{1-\tanh^2\{d/(2L_\mathrm{a})\}}\right]$$

$$= \frac{4D_\mathrm{a}\sinh(d/L_\mathrm{a})}{(\mu_\mathrm{n}+\mu_\mathrm{p})}\frac{1}{\sqrt{1-\frac{1}{4}\tanh^2(d/L_\mathrm{a})}}\times$$

$$\tan^{-1}\left[\frac{2\tanh\left(\dfrac{d}{2L_\mathrm{a}}\right)\sqrt{1-\frac{1}{4}\tanh^2(d/L_\mathrm{a})}}{\mathrm{sech}^2\{d/(2L_\mathrm{a})\}}\right]$$

$$= \frac{4D_\mathrm{a}\sinh(d/L_\mathrm{a})}{(\mu_\mathrm{n}+\mu_\mathrm{p})}\frac{1}{\sqrt{1-\frac{1}{4}\tanh^2(d/L_\mathrm{a})}}\times$$

$$\tan^{-1}\left[2\sinh\left(\frac{d}{2L_\mathrm{a}}\right)\cosh\left(\frac{d}{2L_\mathrm{a}}\right)\sqrt{1-\tfrac{1}{4}\tanh^2(d/L_\mathrm{a})}\right]$$

$$= \frac{4D_\mathrm{a}\sinh(d/L_\mathrm{a})}{(\mu_\mathrm{n}+\mu_\mathrm{p})}\frac{1}{\sqrt{1-\frac{1}{4}\tanh^2(d/L_\mathrm{a})}}\times$$

$$\tan^{-1}\left[\sinh\left(\frac{d}{L_\mathrm{a}}\right)\sqrt{1-\tfrac{1}{4}\tanh^2(d/L_\mathrm{a})}\right] \tag{A4.5.8}$$

But

$$\frac{4D_\mathrm{a}}{(\mu_\mathrm{n}+\mu_\mathrm{p})} = \frac{4\times\dfrac{2D_\mathrm{n}D_\mathrm{p}}{D_\mathrm{n}+D_\mathrm{p}}}{(\mu_\mathrm{n}+\mu_\mathrm{p})} = \frac{8\left(\dfrac{kT}{q}\right)\times\dfrac{\mu_\mathrm{n}\mu_\mathrm{p}}{\mu_\mathrm{n}+\mu_\mathrm{p}}}{(\mu_\mathrm{n}+\mu_\mathrm{p})}$$

$$= 8\left(\frac{kT}{q}\right)\times\frac{\mu_\mathrm{n}\mu_\mathrm{p}}{(\mu_\mathrm{n}+\mu_\mathrm{p})^2} = 8\left(\frac{kT}{q}\right)\times\frac{\mu_\mathrm{n}\mu_\mathrm{p}}{(3\mu_\mathrm{p}+\mu_\mathrm{p})^2}$$

$$= 8\left(\frac{kT}{q}\right)\times\frac{\mu_\mathrm{n}\mu_\mathrm{p}}{(4\mu_\mathrm{p})^2} = \frac{1}{2}\frac{\mu_\mathrm{n}\mu_\mathrm{p}}{\mu_\mathrm{p}^2}\left(\frac{kT}{q}\right) = \frac{1}{2}\left(\frac{\mu_\mathrm{n}}{\mu_\mathrm{p}}\right)\left(\frac{kT}{q}\right)$$

$$= \frac{3}{2}\left(\frac{kT}{q}\right) \tag{A4.5.9}$$

Substituting for $(4D_a)/(\mu_n + \mu_p)$ from Eq. (A4.5.9) into Eq. (A4.5.8) we get the first term of Eq. (A4.5.1) as

$$= \frac{3}{2}\frac{kT}{q}\frac{\sinh(d/L_a)}{\sqrt{1-\frac{1}{4}\tanh^2(d/L_a)}}\tan^{-1}\left[\sinh\left(\frac{d}{L_a}\right)\sqrt{1-\tfrac{1}{4}\tanh^2(d/L_a)}\right] \tag{A4.5.10}$$

Now let us concentrate on the second term of Eq. (A4.5.1). This term is

$$= -\frac{kT}{2q}\int_{-d}^{+d}\frac{\dfrac{\tau_{\mathrm{HL}}J}{2qL_a}\cdot \mathrm{d}\left\{\dfrac{\cosh(x/L_a)}{\sinh(d/L_a)} - \dfrac{1}{2}\dfrac{\sinh(x/L_a)}{\cosh(d/L_a)}\right\}}{\dfrac{\tau_{\mathrm{HL}}J}{2qL_a}\cdot\left\{\dfrac{\cosh(x/L_a)}{\sinh(d/L_a)} - \dfrac{1}{2}\dfrac{\sinh(x/L_a)}{\cosh(d/L_a)}\right\}}$$

$$= -\frac{kT}{2q}\left[\ln\left\{\frac{\cosh(x/L_a)}{\sinh(d/L_a)} - \frac{1}{2}\frac{\sinh(x/L_a)}{\cosh(d/L_a)}\right\}\right]_{-d}^{+d}$$

$$= -\frac{kT}{2q}\left[\ln\left\{\frac{\cosh(d/L_a)}{\sinh(d/L_a)} - \frac{1}{2}\frac{\sinh(d/L_a)}{\cosh(d/L_a)}\right\}\right.$$

$$\left. -\ln\left\{\frac{\cosh(d/L_a)}{\sinh(d/L_a)} + \frac{1}{2}\frac{\sinh(d/L_a)}{\cosh(d/L_a)}\right\}\right]$$

$$= -\frac{kT}{2q}\left[\ln\{\coth(d/L_a) - \tfrac{1}{2}\tanh(d/L_a)\}\right.$$

$$\left. -\ln\{\coth(d/L_a) + \tfrac{1}{2}\tanh(d/L_a)\}\right]$$

$$= \frac{kT}{2q}\cdot\ln\left\{\frac{\coth(d/L_a) + \dfrac{1}{2}\tanh(d/L_a)}{\coth(d/L_a) - \dfrac{1}{2}\tanh(d/L_a)}\right\}$$

$$= \frac{kT}{2q}\cdot\ln\left[\frac{\coth(d/L_a)\{1 + \frac{1}{2}\tanh^2(d/L_a)\}}{\coth(d/L_a)\{1 - \frac{1}{2}\tanh^2(d/L_a)\}}\right]$$

$$= \frac{kT}{q}\cdot\frac{1}{2}\ln\left\{\frac{1 + \frac{1}{2}\tanh^2(d/L_a)}{1 - \frac{1}{2}\tanh^2(d/L_a)}\right\} \tag{A4.5.11}$$

Adding together the first and second terms of Eq. (A4.5.1), as given by Eqs. (A4.5.10) and (A4.5.11), we have

$$V_{\mathrm{N}^-} = \frac{3}{2}\left(\frac{kT}{q}\right)\frac{\sinh(d/L_\mathrm{a})}{\sqrt{1-\frac{1}{4}\tanh^2(d/L_\mathrm{a})}}\tan^{-1}\left[\sinh\left(\frac{d}{L_\mathrm{a}}\right)\sqrt{1-\tfrac{1}{4}\tanh^2(d/L_\mathrm{a})}\right]$$

$$+\frac{1}{2}\left(\frac{kT}{q}\right)\ln\left\{\frac{1+\dfrac{1}{2}\tanh^2(d/L_\mathrm{a})}{1-\dfrac{1}{2}\tanh^2(d/L_\mathrm{a})}\right\} \qquad \text{(A4.5.12)}$$

APPENDIX 4.6: DERIVATION OF CURRENT DENSITY EQUATIONS (4.55) AND (4.56)

From eq. (4.51), the injected hole concentration at the anode junction is

$$p(-d) = p_0 \exp\left(\frac{qV_{\mathrm{P}+}}{kT}\right) \qquad \text{(A4.6.1)}$$

where p_0 is the hole density in the N^- base at thermal equilibrium and $V_{\mathrm{P}+}$ is the potential drop across the anode junction. Similarly, the injected electron concentration at the cathode junction is

$$n(+d) = n_0 \exp\left(\frac{qV_{\mathrm{N}+}}{kT}\right) \qquad \text{(A4.6.2)}$$

where n_0 is the electron density in the N^- base at thermal equilibrium and $V_{\mathrm{N}+}$ is the potential drop across the cathode junction.

Multiplying Eqs. (A4.6.1) and (A4.6.2), we obtain

$$p(-d)\times n(+d) = p_0 \exp\left(\frac{qV_{\mathrm{P}+}}{kT}\right)\times n_0 \exp\left(\frac{qV_{\mathrm{N}+}}{kT}\right)$$

$$= p_0 n_0 \exp\left\{\left(\frac{q}{kT}\right)(V_{\mathrm{P}+}+V_{\mathrm{N}+})\right\} \qquad \text{(A4.6.3)}$$

Applying Eq. (4.46) to $x = -d$, we get

$$p(-d) = \frac{\tau_{\mathrm{HL}} J}{2qL_{\mathrm{a}}} \left\{ \frac{\cosh\left(-\dfrac{d}{L_{\mathrm{a}}}\right)}{\sinh\left(\dfrac{d}{L_{\mathrm{a}}}\right)} - \frac{1}{2} \frac{\sinh\left(-\dfrac{d}{L_{\mathrm{a}}}\right)}{\cosh\left(\dfrac{d}{L_{\mathrm{a}}}\right)} \right\}$$

$$= \frac{\tau_{\mathrm{HL}} J}{2qL_{\mathrm{a}}} \left\{ \coth\left(\frac{d}{L_{\mathrm{a}}}\right) + \frac{1}{2} \tanh\left(\frac{d}{L_{\mathrm{a}}}\right) \right\} \tag{A4.6.4}$$

Similarly, applying Eq. (4.46) to $x = +d$, we obtain

$$n(+d) = \frac{\tau_{\mathrm{HL}} J}{2qL_{\mathrm{a}}} \left\{ \frac{\cosh\left(\dfrac{d}{L_{\mathrm{a}}}\right)}{\sinh\left(\dfrac{d}{L_{\mathrm{a}}}\right)} - \frac{1}{2} \frac{\sinh\left(\dfrac{d}{L_{\mathrm{a}}}\right)}{\cosh\left(\dfrac{d}{L_{\mathrm{a}}}\right)} \right\}$$

$$= \frac{\tau_{\mathrm{HL}} J}{2qL_{\mathrm{a}}} \left\{ \coth\left(\frac{d}{L_{\mathrm{a}}}\right) - \frac{1}{2} \tanh\left(\frac{d}{L_{\mathrm{a}}}\right) \right\} \tag{A4.6.5}$$

Multiplying Eqs. (A4.6.4) and (A4.6.5), we get

$$p(-d) \times n(+d) = \left(\frac{\tau_{\mathrm{HL}} J}{2qL_{\mathrm{a}}} \right)^2 \left\{ \coth^2\left(\frac{d}{L_{\mathrm{a}}}\right) - \frac{1}{4} \tanh^2\left(\frac{d}{L_{\mathrm{a}}}\right) \right\}$$

$$= \left(\frac{\tau_{\mathrm{HL}} J}{2qL_{\mathrm{a}}} \right)^2 \left\{ \frac{1}{\tanh^2\left(\dfrac{d}{L_{\mathrm{a}}}\right)} - \frac{1}{4} \tanh^2\left(\frac{d}{L_{\mathrm{a}}}\right) \right\}$$

$$= \left(\frac{\tau_{\mathrm{HL}} J}{2qL_{\mathrm{a}}} \right)^2 \frac{\left\{ 1 - \dfrac{1}{4} \tanh^4\left(\dfrac{d}{L_{\mathrm{a}}}\right) \right\}}{\tanh^2\left(\dfrac{d}{L_{\mathrm{a}}}\right)}$$

$$= \left(\frac{\tau_{\mathrm{HL}} J}{2qL_{\mathrm{a}}} \right)^2 \frac{\left\{ 1 - 0.25 \tanh^4\left(\dfrac{d}{L_{\mathrm{a}}}\right) \right\}}{\tanh^2\left(\dfrac{d}{L_{\mathrm{a}}}\right)} \tag{A4.6.6}$$

Combining Eqs. (A4.6.3) and (A4.6.6), we obtain

$$p_0 n_o \exp\left\{\left(\frac{q}{kT}\right)(V_{\mathrm{P}+} + V_{\mathrm{N}+})\right\} = \left(\frac{\tau_{\mathrm{HL}} J}{2qL_{\mathrm{a}}}\right)^2 \frac{\left\{1 - 0.25\tanh^4\left(\frac{d}{L_{\mathrm{a}}}\right)\right\}}{\tanh^2\left(\frac{d}{L_{\mathrm{a}}}\right)} \tag{A4.6.7}$$

Taking the square root of both sides and putting $p_0 n_0 = n_{\mathrm{i}}^2$, we get

$$n_i \exp\left\{\left(\frac{q}{2kT}\right)(V_{\mathrm{P}+} + V_{\mathrm{N}+})\right\} = \left(\frac{\tau_{\mathrm{HL}} J}{2qL_{\mathrm{a}}}\right) \frac{\sqrt{1 - 0.25\tanh^4\left(\frac{d}{L_{\mathrm{a}}}\right)}}{\tanh\left(\frac{d}{L_{\mathrm{a}}}\right)} \tag{A4.6.8}$$

This gives the current density

$$J = \frac{2qL_{\mathrm{a}} n_i}{\tau_{\mathrm{HL}}} \frac{\tanh\left(\frac{d}{L_{\mathrm{a}}}\right)}{\sqrt{1 - 0.25\tanh^4\left(\frac{d}{L_{\mathrm{a}}}\right)}} \exp\left\{\left(\frac{q}{2kT}\right)(V_{\mathrm{P}+} + V_{\mathrm{N}+})\right\} \tag{A4.6.9}$$

Since the applied voltage V_{a} = voltage drop across the anode junction ($V_{\mathrm{P}+}$) + voltage drop in the middle N^- base ($V_{\mathrm{N}-}$) + voltage drop across the cathode junction (V_{N^+}), we can write

$$V_{\mathrm{a}} = V_{\mathrm{P}+} + V_{\mathrm{N}-} + V_{\mathrm{N}+} \tag{A4.6.10}$$

giving

$$V_{\mathrm{P}+} + V_{\mathrm{N}+} = V_{\mathrm{a}} - V_{\mathrm{N}-} \tag{A4.6.11}$$

so that

$$\exp(V_{\mathrm{P}+} + V_{\mathrm{N}+}) = \exp(V_{\mathrm{a}} - V_{\mathrm{N}-}) = \exp V_{\mathrm{a}} - \exp V_{\mathrm{N}-} \tag{A4.6.12}$$

Furthermore,

$$\frac{2qL_{\mathrm{a}} n_i}{\tau_{\mathrm{HL}}} = \frac{2qL_{\mathrm{a}}{}^2 n_i}{\tau_{\mathrm{HL}} L_{\mathrm{a}}} = \frac{2qD_{\mathrm{a}}\tau_{\mathrm{a}} n_i}{\tau_{\mathrm{HL}} L_{\mathrm{a}}} = \frac{2qD_{\mathrm{a}} n_i}{d}\left(\frac{d}{L_{\mathrm{a}}}\right) \tag{A4.6.13}$$

where high-level lifetime τ_{HL} has been taken to be equal to ambipolar lifetime τ_{a}. Equations (A4.6.9), (A4.6.12), and (A4.6.13) yield

$$J = \frac{2qD_{\mathrm{a}}n_i}{d}\left(\frac{d}{L_{\mathrm{a}}}\right)\frac{\tanh\left(\frac{d}{L_{\mathrm{a}}}\right)}{\sqrt{1-0.25\tanh^4\left(\frac{d}{L_{\mathrm{a}}}\right)}}\exp\left\{\left(\frac{qV_{\mathrm{a}}}{2kT}\right)-\left(\frac{qV_{\mathrm{N-}}}{2kT}\right)\right\}$$

$$= \frac{2qD_{\mathrm{a}}n_i}{d}F\left(\frac{d}{L_{\mathrm{a}}}\right)\exp\left\{-\left(\frac{qV_{\mathrm{N-}}}{2kT}\right)\right\} \tag{A4.6.14}$$

where

$$F\left(\frac{d}{L_{\mathrm{a}}}\right) = \left(\frac{d}{L_{\mathrm{a}}}\right)\frac{\tanh\left(\frac{d}{L_{\mathrm{a}}}\right)}{\sqrt{1-0.25\tanh^4\left(\frac{d}{L_{\mathrm{a}}}\right)}}\exp\left(\frac{qV_{\mathrm{a}}}{2kT}\right) \tag{A4.6.15}$$

APPENDIX 4.7: TRANSISTOR TERMINAL CURRENTS [EQS. (4.57) AND (4.58)]

The excess hole concentration (Δp) as a function of distance (x) in the base is given by the solution of the diffusion equation

$$\frac{d^2\Delta p}{dx^2} = \frac{\Delta p}{L_{\mathrm{B}}^2} \tag{A4.7.1}$$

where L_{B} is the diffusion length of holes in the base region. The solution of this equation is expressed as

$$\Delta p = K_1\exp\left(\frac{x}{L_{\mathrm{B}}}\right) + K_2\exp\left(-\frac{x}{L_{\mathrm{B}}}\right) \tag{A4.7.2}$$

where K_1 and K_2 are constants determined by the boundary conditions

$$\Delta p|_{x=0} = \Delta p_{\mathrm{E}} \tag{A4.7.3}$$

$$\Delta p|_{x=W} = \Delta p_{\mathrm{C}} \tag{A4.7.4}$$

where Δp_{E} and Δp_{C} are the hole concentrations on the emitter and collector sides, respectively, and W is the base width. Applying Eqs. (A4.7.3) and (A4.7.4) to Eq. (A4.7.2), we get

$$K_1 + K_2 = \Delta p_{\mathrm{E}} \tag{A4.7.5}$$

$$K_1 \exp\left(\frac{W}{L_{\mathrm{B}}}\right) + K_2 \exp\left(-\frac{W}{L_{\mathrm{B}}}\right) = \Delta p_{\mathrm{C}} \tag{A4.7.6}$$

After solving these equations, we have

$$K_1 = \frac{\Delta p_{\mathrm{C}} - \Delta p_{\mathrm{E}} \exp(-W/L_{\mathrm{B}})}{\exp(W/L_{\mathrm{B}}) - \exp(-W/L_{\mathrm{B}})} \tag{A4.7.7}$$

$$K_2 = \frac{\Delta p_{\mathrm{E}} \exp(W/L_{\mathrm{B}}) - \Delta p_{\mathrm{C}}}{\exp(W/L_{\mathrm{B}}) - \exp(-W/L_{\mathrm{B}})} \tag{A4.7.8}$$

Substituting for K_1 and K_2 from Eqs. (A4.7.7) and (A4.7.8) into Eq. (A4.7.2), we obtain

$$\Delta p = \frac{\Delta p_{\mathrm{C}} - \Delta p_{\mathrm{E}} \exp(-W/L_{\mathrm{B}})}{\exp(W/L_{\mathrm{B}}) - \exp(-W/L_{\mathrm{B}})} \exp\left(\frac{x}{L_{\mathrm{B}}}\right) + \frac{\Delta p_{\mathrm{E}} \exp(W/L_{\mathrm{B}}) - \Delta p_{\mathrm{C}}}{\exp(W/L_{\mathrm{B}}) - \exp(-W/L_{\mathrm{B}})} \exp\left(-\frac{x}{L_{\mathrm{B}}}\right) \tag{A4.7.9}$$

After solving for the excess hole distribution in the base, the emitter and collector currents are determined from the gradient of the hole concentration at the emitter and collector depletion region edges.

Now the hole current is expressed as

$$I_{\mathrm{p}} = -qAD_{\mathrm{p}} \frac{d\Delta p}{dx} = -qAD_{\mathrm{B}} \frac{d\Delta p}{dx} \tag{A4.7.10}$$

$$\begin{aligned} I_{\mathrm{p}} &= -qAD_{\mathrm{B}} \frac{d\Delta p}{dx} \\ &= -qAD_{\mathrm{B}} \frac{d}{dx} \left\{ \frac{\Delta p_{\mathrm{C}} - \Delta p_{\mathrm{E}} \exp(-W/L_{\mathrm{B}})}{\exp(W/L_{\mathrm{B}}) - \exp(-W/L_{\mathrm{B}})} \exp\left(\frac{x}{L_{\mathrm{B}}}\right) \right. \\ &\quad \left. + \frac{\Delta p_{\mathrm{E}} \exp(W/L_{\mathrm{B}}) - \Delta p_{\mathrm{C}}}{\exp(W/L_{\mathrm{B}}) - \exp(-W/L_{\mathrm{B}})} \exp\left(-\frac{x}{L_{\mathrm{B}}}\right) \right\} \end{aligned}$$

$$= -qAD_{\mathrm{B}}\left\{\frac{\Delta p_{\mathrm{C}} - \Delta p_{\mathrm{E}}\exp(-W/L_{\mathrm{B}})}{\exp(W/L_{\mathrm{B}}) - \exp(-W/L_{\mathrm{B}})}\exp\left(\frac{x}{L_{\mathrm{B}}}\right)\left(\frac{1}{L_{\mathrm{B}}}\right)\right.$$

$$\left. + \frac{\Delta p_{\mathrm{E}}\exp(W/L_{\mathrm{B}}) - \Delta p_{\mathrm{C}}}{\exp(W/L_{\mathrm{B}}) - \exp(-W/L_{\mathrm{B}})}\exp\left(-\frac{x}{L_{\mathrm{B}}}\right)\left(-\frac{1}{L_{\mathrm{B}}}\right)\right\}$$

$$= \frac{-qAD_{\mathrm{B}}}{L_{\mathrm{B}}}\left\{\frac{\Delta p_{\mathrm{C}}\exp(x/L_{\mathrm{B}}) - \Delta p_{\mathrm{E}}\exp(-W/L_{\mathrm{B}})\exp(x/L_{\mathrm{B}})}{\exp(W/L_{\mathrm{B}}) - \exp(-W/L_{\mathrm{B}})}\right.$$

$$\left. - \frac{\Delta p_{\mathrm{E}}\exp(W/L_{\mathrm{B}})\exp(-x/L_{\mathrm{B}}) - \Delta p_{\mathrm{C}}\exp(-x/L_{\mathrm{B}})}{\exp(W/L_{\mathrm{B}}) - \exp(-W/L_{\mathrm{B}})}\right\}$$

$$= \frac{-qAD_{\mathrm{B}}}{L_{\mathrm{B}}} \times \frac{1}{\exp(W/L_{\mathrm{B}}) - \exp(-W/L_{\mathrm{B}})}$$

$$\times\{\Delta p_{\mathrm{C}}\exp(x/L_{\mathrm{B}}) - \Delta p_{\mathrm{E}}\exp(-W/L_{\mathrm{B}})\exp(x/L_{\mathrm{B}})$$

$$+\Delta p_{\mathrm{E}}\exp(W/L_{\mathrm{B}})\exp(-x/L_{\mathrm{B}}) - \Delta p_{\mathrm{C}}\exp(-x/L_{\mathrm{B}})\}$$

$$= \frac{-qAD_{\mathrm{B}}}{L_{\mathrm{B}}} \times \frac{1}{\exp(W/L_{\mathrm{B}}) - \exp(-W/L_{\mathrm{B}})}$$

$$\times\left[\begin{array}{l}\Delta p_{\mathrm{E}}\{\exp(-W/L_{\mathrm{B}})\exp(x/L_{\mathrm{B}}) + \exp(W/L_{\mathrm{B}})\exp(-x/L_{\mathrm{B}})\}\\ -\Delta p_{\mathrm{C}}\{\exp(x/L_{\mathrm{B}}) - \exp(-x/L_{\mathrm{B}})\}\end{array}\right]$$

$$= \frac{qAD_{\mathrm{B}}}{2L_{\mathrm{B}}\sinh(W/L_{\mathrm{B}})} \times [\Delta p_{\mathrm{E}}\{\exp(-W/L_{\mathrm{B}})\exp(x/L_{\mathrm{B}})$$

$$+\exp(W/L_{\mathrm{B}})\exp(-x/L_{\mathrm{B}})\} - \Delta p_{\mathrm{C}}\{2\cosh(x/L_{\mathrm{B}})\}] \qquad \text{(A4.7.11)}$$

where the formula $\exp(z) - \exp(-z) = 2\sinh(z)$ has been applied.

$$I_{\mathrm{p}}\big|_{x=0} = I_{\mathrm{pE}}$$

$$= \frac{qAD_{\mathrm{B}}}{2L_{\mathrm{B}}\sinh(W/L_{\mathrm{B}})} \times [\Delta p_{\mathrm{E}}\{\exp(-W/L_{\mathrm{B}}) + \exp(W/L_{\mathrm{B}})\} - 2\Delta p_{\mathrm{C}}]$$

$$= \frac{qAD_{\mathrm{B}}}{2L_{\mathrm{B}}\sinh(W/L_{\mathrm{B}})} \times \{\Delta p_{\mathrm{E}} \times 2\cosh(W/L_{\mathrm{B}}) - 2\Delta p_{\mathrm{C}}\}$$

$$= \left(\frac{qAD_{\mathrm{B}}}{L_{\mathrm{B}}}\right)\{\coth(W/L_{\mathrm{B}})\Delta p_{\mathrm{E}} - \Delta p_{\mathrm{C}}\,\mathrm{cosech}(W/L_{\mathrm{B}})\} \qquad \text{(A4.7.12)}$$

where I_{pE} is the hole component of emitter current.

At $x = W$,

$$I_p\big|_{x=W} = I_{pC} = \frac{qAD_B}{2L_B \sinh(W/L_B)} \times \begin{bmatrix} \Delta p_E\{\exp(-W/L_B)\exp(W/L_B) \\ +\exp(W/L_B)\exp(-W/L_B)\} \\ -\Delta p_C\{2\cosh(W/L_B)\} \end{bmatrix}$$

$$= \frac{qAD_B}{2L_B \sinh(W/L_B)} \times [2\Delta p_E - 2\Delta p_C \cosh(W/L_B)]$$

$$= \left(\frac{qAD_B}{L_B}\right)\frac{1}{\sinh(W/L_B)}\{\Delta p_E - \Delta p_C \cosh(W/L_B)\} \quad \text{(A4.7.13)}$$

where I_{pC} is the hole component of collector current.

Now, the excess hole concentration Δp_E at the edge of the emitter–base depletion region on the base side is given by

$$\Delta p_E = p(0) - p_B = p_B\left\{\exp\left(\frac{qV_{EB}}{kT}\right) - 1\right\} \quad \text{(A4.7.14)}$$

where p_B is the equilibrium hole density in the base and V_{EB} is the emitter–base bias. Similarly, the excess hole concentration Δp_C at the edge of the collector–base depletion region on the base side is

$$\Delta p_C = p(W) - p_B = p_B\left\{\exp\left(\frac{qV_{CB}}{kT}\right) - 1\right\} \quad \text{(A4.7.15)}$$

where V_{CB} is the collector–base bias.

Substituting for Δp_E from Eq. (A4.7.14) and Δp_C from Eq. (A4.7.15) into Eqs. (A4.7.12) and (A4.7.13) respectively, we obtain

$$I_{pE} = \left(\frac{qAD_B}{L_B}\right)\left[\coth(W/L_B)\,p_B\left\{\exp\left(\frac{qV_{EB}}{kT}\right) - 1\right\} - p_B\left\{\exp\left(\frac{qV_{CB}}{kT}\right) - 1\right\}\operatorname{cosech}(W/L_B)\right]$$

$$= \left(\frac{qAD_B p_B}{L_B}\right)\coth(W/L_B)\left[\left\{\exp\left(\frac{qV_{EB}}{kT}\right) - 1\right\} - \operatorname{sech}(W/L_B)\left\{\exp\left(\frac{qV_{CB}}{kT}\right) - 1\right\}\right] \quad \text{(A4.7.16)}$$

and

$$I_{pC} = \left(\frac{qAD_B}{L_B}\right)\left[p_B\left\{\exp\left(\frac{qV_{EB}}{kT}\right) - 1\right\} - p_B\left\{\exp\left(\frac{qV_{CB}}{kT}\right) - 1\right\}\cosh(W/L_B)\right]$$

$$= \left(\frac{qAD_B p_B}{L_B}\right)\frac{1}{\sinh(W/L_B)}\left[\left\{\exp\left(\frac{qV_{EB}}{kT}\right) - 1\right\}\right.$$

$$\left. -\cosh(W/L_B)\left\{\exp\left(\frac{qV_{CB}}{kT}\right) - 1\right\}\right] \qquad \text{(A4.7.17)}$$

The electron current crossing from the base into the emitter at the depletion region edge on the emitter side $x = -x_E$ is

$$I_{nE}|_{x=-x_E} = \left(\frac{qAD_E \Delta n_E}{L_E}\right) \qquad \text{(A4.7.18)}$$

where D_{nE} and L_E are, respectively, the electron diffusion coefficient and diffusion length in the emitter and Δn_E is the excess electron concentration in the emitter. But

$$\Delta n_E = n(-x_E) - n_E = n_E\left\{\exp\left(\frac{qV_{EB}}{kT}\right) - 1\right\} \qquad \text{(A4.7.19)}$$

From Eqs. (A4.7.18) and (A4.7.19), we obtain

$$I_{nE}|_{x=-x_E} = \left(\frac{qAD_E n_E}{L_E}\right)\left\{\exp\left(\frac{qV_{EB}}{kT}\right) - 1\right\} \qquad \text{(A4.7.20)}$$

Proceeding on similar lines, the electron current flowing from the base into the collector across the depletion region edge on the collector side $x = x_C$ is

$$I_{nC}|_{x=x_C} = -\left(\frac{qAD_C \Delta n_C}{L_C}\right) \qquad \text{(A4.7.21)}$$

where D_{nC} and L_C are, respectively, the electron diffusion coefficient and diffusion length in the collector and Δn_C is the excess electron concentration in the collector. But

$$\Delta n_C = n(x_C) - n_C = n_C\left\{\exp\left(\frac{qV_{CB}}{kT}\right) - 1\right\} \qquad \text{(A4.7.22)}$$

Equations (A4.7.21) and (A4.7.22) give

$$I_{nC}|_{x=x_C} = -\left(\frac{qAD_C n_C}{L_C}\right)\left\{\exp\left(\frac{qV_{CB}}{kT}\right) - 1\right\} \qquad \text{(A4.7.23)}$$

Addition of Equations (A4.7.16) and (A4.7.20) gives the total emitter current

$$I_E = I_{pE} + I_{nE} = \left(\frac{qAD_B p_B}{L_B}\right)\coth(W/L_B)$$

$$\times\left[\left\{\exp\left(\frac{qV_{EB}}{kT}\right)-1\right\} - \operatorname{sech}(W/L_B)\left\{\exp\left(\frac{qV_{CB}}{kT}\right)-1\right\}\right]$$

$$+\left(\frac{qAD_E n_E}{l_E}\right)\left\{\exp\left(\frac{qV_{EB}}{kT}\right)-1\right\} \tag{A4.7.24}$$

Likewise, the total collector current is obtained by the addition of Eqs. (A4.7.17) and (A4.7.23):

$$I_C = I_{pC} + I_{nC} = \left(\frac{qAD_B p_B}{L_B}\right)\frac{1}{\sinh(W/L_B)}\left[\left\{\exp\left(\frac{qV_{EB}}{kT}\right)-1\right\}\right.$$

$$\left. -\cosh(W/L_B)\left\{\exp\left(\frac{qV_{CB}}{kT}\right)-1\right\}\right]$$

$$-\left(\frac{qAD_C n_C}{L_C}\right)\left\{\exp\left(\frac{qV_{CB}}{kT}\right)-1\right\} \tag{A4.7.25}$$

APPENDIX 4.8: COMMON-BASE CURRENT GAIN α_T [EQ. (4.63)]

Let us consider the emitter base PN diode of a PNP transistor. The differential equation describing the injected excess hole concentration in the N region is

$$D_p\frac{d^2\hat{p}}{dx^2} = \frac{\hat{p}}{\tau_p} \tag{A4.8.1}$$

where D_p is the hole diffusion coefficient, τ_p is the hole lifetime, and the general hole concentration p is replaced by the excess hole concentration $\hat{p}$ since $p = \bar{p} + \hat{p}$ and the thermal equilibrium hole concentration $\bar{p}$ does not change with distance. Then Eq. (A4.8.1) may be rewritten as

$$\frac{d^2\hat{p}}{dx^2} = \frac{\hat{p}}{D_p\tau_p} \tag{A4.8.2}$$

This is a simple differential equation describing the injected excess hole concentration in the N region. The solution to this equation is the sum of two exponentials:

$$\hat{p} = C_1 \exp\left(\frac{x}{\sqrt{D_p \tau_p}}\right) + C_2 \exp\left(-\frac{x}{\sqrt{D_p \tau_p}}\right) \tag{A4.8.3}$$

where C_1 and C_2 are constants to be determined from the boundary conditions. The factor $\sqrt{(D_p \tau_p)}$ has the dimensions of length and is called the *diffusion length* L_p. (At $x = L_p$, the excess hole concentration is $\hat{p}_0/e$, and therefore L_p is the average range of the excess holes in the N region). In terms of the hole diffusion length $L_p = \sqrt{(D_p \tau_p)}$, Eq. (A4.8.3) is altered to the form

$$\hat{p} = C_1 \exp\left(\frac{x}{L_p}\right) + C_2 \exp\left(-\frac{x}{L_p}\right) \tag{A4.8.4}$$

One of the boundary conditions is that $\hat{p} = 0$ at $x = W =$ base width. Then

$$0 = C_1 \exp\left(\frac{W}{L_p}\right) + C_2 \exp\left(-\frac{W}{L_p}\right) \tag{A4.8.5}$$

from which

$$C_1 = -\frac{C_2 \exp(-W/L_p)}{\exp(W/L_p)} = -C_2 \exp(-2W/L_p) \tag{A4.8.6}$$

The second boundary condition is as follows: at $x = 0$, $\hat{p} = \hat{p}_0 =$ excess concentration of holes at the boundary of the depletion layer, that is, $\hat{p} = \hat{p}_0$ at $x = 0$. Therefore, Eq. (A4.8.4) gives

$$\hat{p}_0 = C_1 + C_2 \tag{A4.8.7}$$

Substituting the value of C_1 from Eq. (A4.8.6) into Eq. (A4.8.7), we get

$$\hat{p}_0 = -C_2 \exp(-2W/L_p) + C_2 \tag{A4.8.8}$$

yielding

$$C_2 = \frac{\hat{p}_0}{1 - \exp(-2W/L_p)} \tag{A4.8.9}$$

From Eqs. (A4.8.6) and (A4.8.9), we have

$$C_1 = -\frac{\hat{p}_0}{1 - \exp(-2W/L_p)} \exp(-2W/L_p) \tag{A4.8.10}$$

Putting the values of constants C_1 and C_2 from Eqs. (A4.8.10) and (A4.8.9) into Eq. (A4.8.4), we obtain

$$\hat{p} = -\frac{\hat{p}_0}{1-\exp(-2W/L_p)}\exp(-2W/L_p)\exp\left(\frac{x}{L_p}\right)$$

$$+\frac{\hat{p}_0}{1-\exp(-2W/L_p)}\exp\left(-\frac{x}{L_p}\right) \qquad \text{(A4.8.11)}$$

which is rearranged as

$$\hat{p} = \hat{p}_0\left\{\frac{-\exp(x/L_p)\exp(-2W/L_p)+\exp(-x/L_p)}{1-\exp(-2W/L_p)}\right\} \qquad \text{(A4.8.12)}$$

which represents the hole distribution between the depletion layer and the ohmic contact.

Now, the base transport factor α_T is expressed as

$$\alpha_T = \frac{\text{Hole current at } W}{\text{Hole current at emitter}} = \frac{\left.\frac{d\hat{p}}{dx}\right|_{x=W}}{\left.\frac{d\hat{p}}{dx}\right|_{\text{at emitter}}} \qquad \text{(A4.8.13)}$$

Differentiating Eq. (A4.8.12),

$$\frac{d\hat{p}}{dx} = -\frac{\hat{p}_0}{\{1-\exp(2W/L_p)\}}$$

$$\times\left\{-\exp(-2W/L_p)\times\exp(x/L_p)\times\frac{1}{L_p}+\exp(-x/L_p)\times-\frac{1}{L_p}\right\}$$

$$\frac{d\hat{p}}{dx} = \frac{\hat{p}_0}{L_p\{1-\exp(2W/L_p)\}}\{\exp(-2W/L_p)\times\exp(x/L_p)+\exp(-x/L_p)\}$$

(A4.8.14)

At $x = 0$,

$$\left.\frac{d\hat{p}}{dx}\right|_{x=0} = \frac{\hat{p}_0}{L_p\{1-\exp(2W/L_p)\}}\{\exp(-2W/L_p)\times 1+1\} \qquad \text{(A4.8.15)}$$

At $x = W$,

$$\left.\frac{d\hat{p}}{dx}\right|_{x=W} = \frac{\hat{p}_0}{L_p\{1-\exp(2W/L_p)\}} \times\{\exp(-2W/L_p)\times\exp(W/L_p)+\exp(-W/L_p)\}$$

$$= \frac{\hat{p}_0}{L_p\{1-\exp(2W/L_p)\}}\{\exp(-W/L_p)+\exp(-W/L_p)\}$$

$$= \frac{\hat{p}_0}{L_p\{1-\exp(2W/L_p)\}}\times 2\exp(-W/L_p) \qquad \text{(A4.8.16)}$$

Inserting the values of $d\hat{p}/dx$ at $x=0$ and $x=W$ into Eq. (A4.8.13), we obtain

$$\alpha_T = \frac{2\exp(-W/L_p)}{1+\exp(-2W/L_p)} = \frac{1}{\dfrac{1+\exp(-2W/L_p)}{2\exp(-W/L_p)}}$$

$$= \frac{1}{\dfrac{\exp(W/L_p)+\exp(-W/L_p)}{2}} = \frac{1}{\cosh(W/L_p)} \qquad \text{(A4.8.17)}$$

Now

$$\cosh x = 1+\frac{x^2}{2!}+\frac{x^4}{4!}+\cdots \qquad \text{(A4.8.18)}$$

Therefore,

$$\frac{1}{\cosh x} = (\cosh x)^{-1} = \left(1+\frac{x^2}{2!}+\frac{x^4}{4!}+\cdots\right)^{-1} \cong 1-\frac{x^2}{2!} \qquad \text{(A4.8.19)}$$

because

$$(1+x)^{-1} = 1-x+x^2-x^3+x^4-\cdots \cong 1-x \qquad \text{(A4.8.20)}$$

From Eqs. (A4.8.17) and (A4.8.19), we get

$$\alpha_T = 1-\frac{W^2}{2L_p^{\,2}} \qquad \text{(A4.8.21)}$$

L_p may be replaced by L_B because they represent the same quantity. Then

$$\alpha_T = 1-\frac{W^2}{2L_B^{\,2}} \qquad \text{(A4.8.22)}$$

5

PHYSICS AND MODELING OF IGBT

An IGBT model must be able to simulate the IGBT or the circuit using this component with adequate accuracy in a reasonable computation time. IGBT modeling can be pursued along three different approaches: (i) *Analytical Device and Circuit Simulation*: The IGBT is considered as a special connection of discrete components (PIN diode and MOSFET or MOSFET and bipolar transistor). Appropriate parameters of these components are applied to calculate the output characteristics. As a first step, the approach provides physical understanding of device behavior. It is useful for simulating device operation but not for construction of a circuit using the devices to emulate IGBT curves. To study the IGBT as a circuit component, it is necessary to extend the basic device model to include the device–circuit interactions for IGBTs under general loading conditions. Role of snubber protection circuit must also be analyzed. (ii) *Numerical Device Simulation*: Physical operation of the device is described in terms of drift and diffusion equations. These are called *isothermal simulations*. Together with the heat flow equation, *electrothermal*, *nonisothermal*, or *thermodynamic simulations* are performed. Exact knowledge of device structural parameters, doping profiles, carrier lifetimes, and so on, is required along with a great deal of computational power. The above-mentioned equations are solved in one, two or three dimensions. The significant advantage gained is the precision or exactness of the results but the simulations are usually lengthy and time-consuming, necessitating patience and perseverance. (iii) *External Behavior Modeling*: Here we direct

---

The Insulated Gate Bipolar Transistor (IGBT) Theory and Design, By Vinod Kumar Khanna.
ISBN 0-471-23845-7 © 2003 Institute of Electrical and Electronics Engineers

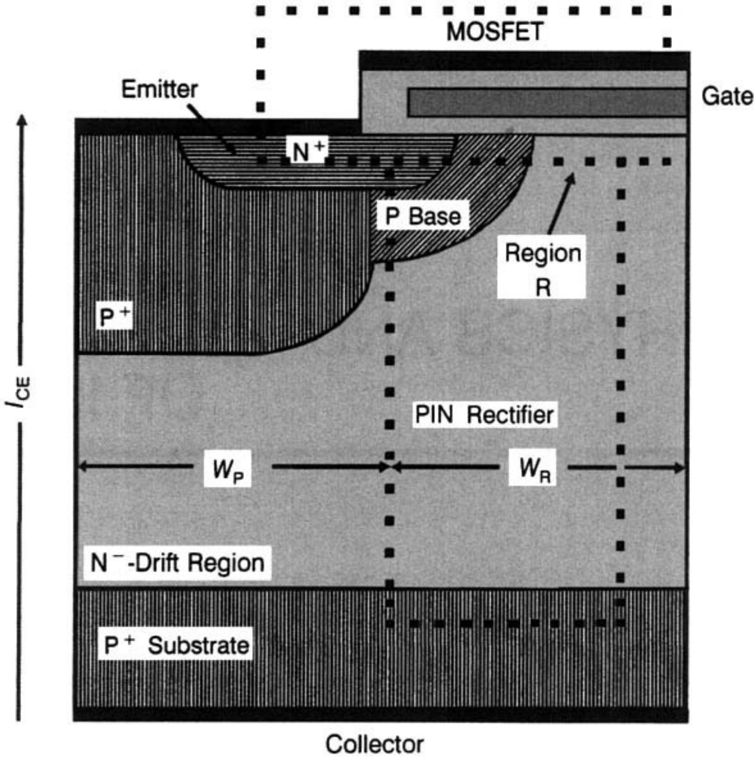


Figure 5.1 Schematic diagram of IGBT as a combination of PIN rectifier and MOSFET.

our attention on only the input and output characteristics of the devices without any concern for its internal structure and related effects. Such a model is built up from steered voltage and current sources. This approach is more suitable for complex circuit calculations.

5.1 PIN RECTIFIER–DMOSFET MODEL OF IGBT

It would be expedient to recall the IGBT equivalent circuits and circuit models presented in Sections 1.5 and 1.7.

5.1.1 Basic Model Formulation

In this model (Fig. 5.1), the IGBT is seen as a series connection of PIN rectifier and MOSFET [1]. The current–voltage characteristics of IGBT are calculated by applying the equations for current conduction through the PIN diode and the MOSFET derived in the preceding chapters. To apply these equations, the common potential of these components, at region R in the structure, is used. The current density J_F and the voltage drop V_F across a PIN rectifier are related as follows [Eq. (4.55)]:

$$J_F = \frac{2qD_a n_i}{d} F\left(\frac{d}{L_a}\right) \exp\left(\frac{qV_F}{2kT}\right) \tag{5.1}$$

The current injected by the P^+ substrate and entering the drift region from the bottom is uniformly distributed in this region across the device cross section over a considerable portion of the drift region. Let Z denote the length perpendicular to the cross section of the linear IGBT cell. The current density J_F is expressed in terms of the current I_{CE} as

$$J_F = \frac{I_{CE}}{W_R Z} \tag{5.2}$$

Combining Eqs. (5.1) and (5. 2), we obtain

$$V_F = \left(\frac{2kT}{q}\right) \ln\left\{\frac{I_{CE} d}{2qW_R Z D_a n_i F(d/L_a)}\right\} \tag{5.3}$$

As this current is constrained to flow through the MOSFET channel, the output current I_{CE} of the MOSFET can be written in terms of forward potential drop across the MOSFET ($V_{F,MOS}$) and the gate-emitter voltage V_{GE} of the IGBT as [Eq. (3.22)]

$$I_{CE} = \frac{\mu_{ox} C_{ox} Z}{L_{Ch}} (V_{GE} - V_{Th}) V_{F,MOS} \tag{5.4}$$

where $V_{F,MOS}$ is used for V_{DS} and L_{Ch} is the channel length. This equation naturally follows by recognizing that in the forward conduction mode, the applied gate–emitter voltage is sufficiently high to make the forward drop across the device low, that is, $V_{GE} - V_{Th} \gg V_{F,MOS}$. Consequently, the MOSFET section of the IGBT is in the linear operating region. Algebraic manipulation of I_{CE} equation gives the voltage drop across the MOSFET portion as

$$V_{F,MOS} = \frac{I_{CE} L_{Ch}}{\mu_{ns} C_{ox} Z (V_{GE} - V_{Th})} \tag{5.5}$$

By adding the potential drops across the PIN rectifier and MOSFET portions, the forward drop of IGBT is obtained as

$$V_{CE} = \left(\frac{2kT}{q}\right) \ln\left\{\frac{I_{CE} d}{2qW_R D_a n_i F(d/L_a)}\right\} + \frac{I_{CE} L_{Ch}}{\mu_{ns} C_{ox} Z (V_{GE} - V_{Th})} \tag{5.6}$$

This equation allows us to calculate the output I_{CE}–V_{CE} characteristics of IGBT at different gate bias voltages V_{GE}. These characteristics show a distinctive *diode knee or kink below* which negligible current flows because there is no injection from P^+ substrate. This indicates the unsuitability of using IGBTs in applications where forward drops below 0.7 V are necessary. Above 0.7 V, the equation predicts an exponential increase of forward current with voltage as in a PIN rectifier. Furthermore, the characteristics are uniformly spaced, but in reality they tend to clump together at higher gate voltages due to channel mobility reduction at high electric fields.

At low values of gate bias, V_{GE}, appreciable voltage drop takes place in the channel and the MOSFET portion of the IGBT limits the current. Therefore collector current saturation occurs at a value

$$I_{CE} = \frac{\mu_{ns} C_{ox} Z}{2L_{Ch}} (V_{GE} - V_{Th})^2 \tag{5.7}$$

The PIN rectifier–DMOSFET model of IGBT is able to provide an explanation of the IGBT behavioral pattern with carrier lifetime variation. This becomes apparent on analyzing the impact of L_a on PIN rectifier characteristics. It also allows the study of the influence of increasing the breakdown voltage through the dependence of PIN rectifier characteristics on N^--base width and doping concentration. The principal shortcoming of the model is that the hole component of collector–emitter current I_{CE} flowing into the P base below the N^+ emitter is neglected. Due consideration of this component is taken in the bipolar transistor–MOSFET model of IGBT, which will be treated in the ensuing section.

5.1.2 Carrier Distribution in IGBT Drift Region in the ON State

The basic model outlined above assumes that the boundary conditions for IGBTs are similar to those for the PIN rectifier in the forward conducting state [1]. However, a major dissimilarity exists because the junction between the N^--drift region and P base of the IGBT is reverse-biased. Thus, the free carrier concentration must fall to zero at this junction. So the conditions prevailing in an IGBT are analogous to the PIN rectifier at high-level injection, with the exception that free carrier density is zero at the junction indicated. By performing a one-dimensional analysis from the P^+-substrate/N^--drift region junction to N^--drift region/P-base junction, the free carrier distribution in IGBT is obtained. This involves solving the continuity equation [Eq. (4.28)] under steady-state conditions (Appendix 5.1). This equation is

$$\frac{d^2p}{dx^2} = \frac{p}{{L_{HL}}^2} \tag{5.8}$$

and the boundary conditions are: $p(0)$ = hole concentration in the N$^-$-drift region at the junction between P$^+$ substrate and N$^-$-drift region $= p_0$, $p(d)$ = hole concentration in the N$^-$-drift region at the junction between N$^-$-drift region and P base = 0. Then the solution is given by (Appendix 5.1)

$$p(x) = \frac{p_0 \sinh\{(d-x)/L_a\}}{\sinh(d/L_a)} \tag{5.9}$$

In this solution, p_0 must be determined. Using the boundary conditions for the hole and electron currents, p_0 is found out. In terms of the collector current density J, these boundary conditions are specified as $J_n(x) = 0$ at $x = 0$ and $J_p(x) = J$ at $x = 0$. Employing these conditions, the equation for p_0 is obtained as

$$p_0 = \frac{JL_a}{2qD_p} \tanh\left(\frac{d}{L_a}\right) \tag{5.10}$$

Thus from Eq. (5.9), the expression for the carrier distribution in the drift region is written as

$$p(x) = \frac{JL_a}{2qD_p} \frac{\sinh\left(\dfrac{d-x}{L_a}\right)}{\cosh(d/L_a)} \tag{5.11}$$

The carrier distribution profile of IGBT is compared with the catenary carrier distribution of PIN rectifier in Fig. 5.2. Near the junction between N$^-$-drift region/P$^+$ substrate, the two carrier distributions are identical. But

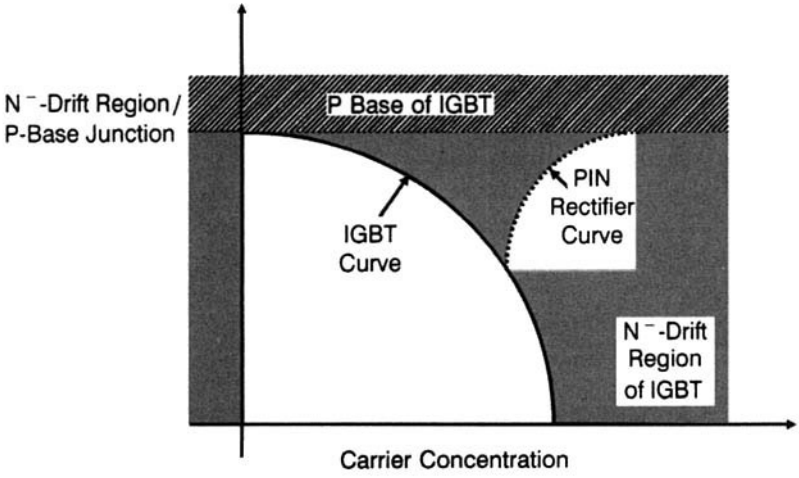


Figure 5.2 Difference between ON-state carrier distribution in the PIN rectifier and IGBT.

at the junction between N$^-$-drift region/P base, they are widely different. At this junction, the carrier distribution in the IGBT falls to zero but that in the PIN rectifier remains high. Naturally therefore, close to this junction, the conductivity modulation effect in the IGBT is small but that in the PIN rectifier is appreciably large. Consequently, the forward voltage of the IGBT is larger than that of the PIN rectifier, and the JFET region in IGBT vastly contributes to this potential drop.

Example 5.1 An IGBT is fabricated with a drift region thickness of 100 μm (excluding the P base and N$^+$ layers). Carrier lifetime in the N$^-$-drift region is 1 μsec. The IGBT is operated at a current density of 200 A/cm^2. During operation, the depletion layer extends into the drift region up to 3 μm. Plot the hole distribution profile in the IGBT.

The IGBT structure is shown in Fig. E5.1.1. Here, $J = 200$ A/cm^2, $d = 100 \times 10^{-4}$ cm, $L_a = \sqrt{(D_a \tau_a)} = \sqrt{(18.34 \times 1 \times 10^{-6})} = 4.283 \times 10^{-3}$ cm. Since $D_p = 12.43$ cm^2/sec, for $x = 0$, we have from Eq. (5.11), $p(x) = \{(200 \times 4.283 \times 10^{-3})/(2 \times 1.6 \times 10^{-19} \times 12.43)\}[\sinh\{(100 - 3) \times 10^{-4} - 0\}/(4.283 \times 10^{-3}]/[\cosh\{(100 - 3) \times 10^{-4} - 0\}/(4.283 \times 10^{-3}] = 2.1 \times 10^{17}$ cm^{-3}. Putting $x = 10$ μm, we get $p(x) = 1.62 \times 10^{17}$ cm^{-3}. Similarly, putting successively $x = 20$, 30, 40, 50, 60, 70, 80, 90, 95, 96.5, 96.8, 97 μm, we get the corresponding values of $p(x)$ that are compiled in Table E5.1.1. The graph showing the carrier distribution is presented in Fig. E5.1.2.

Significance of This Example:. This example illustrates the carrier (hole) distribution in the N$^-$-drift region of an IGBT during the conducting state. This charge distribution determines the conductivity modulation of the N$^-$-drift region and thereby the forward voltage dropped across the IGBT. It is evident that the injected hole concentration has the peak value in the vicinity of the junction between P$^+$ substrate and N$^-$-drift regions. This concentration falls as one recedes away from this junction. So, the portion of the N$^-$-drift region farthest away from this junction contributes the largest part to the forward drop. Furthermore, mainly the operating current density J, the ambipolar diffusion length L_a, and thickness of N$^-$-drift region control the injected hole distribution.

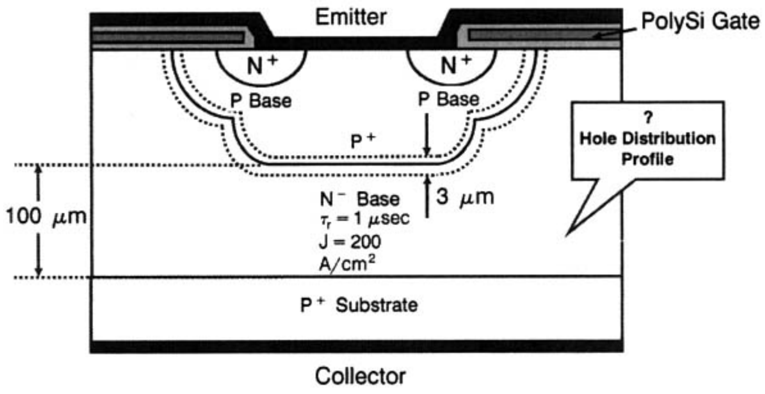


Figure E5.1.1 IGBT for Example 5.1.

Table E5.1.1 Hole Distribution at Different Distances in the N$^-$-Drift Region of IGBT

Serial No.	Distance x (μm)	Hole Concentration $p(x)$, cm^{-3}
1	0	2.1×10^{17}
2	10	1.62×10^{17}
3	20	1.267×10^{17}
4	30	9.86×10^{17}
5	40	7.595×10^{16}
6	50	5.745×10^{16}
7	60	4.192×10^{16}
8	70	2.904×10^{16}
9	80	1.758×10^{16}
10	90	7.084×10^{15}
11	95	2.016×10^{15}
12	96.5	5.04×10^{14}
13	96.8	2.015×10^{14}
14	96.9	1.0075×10^{14}
15	96.99	1.0075×10^{13}

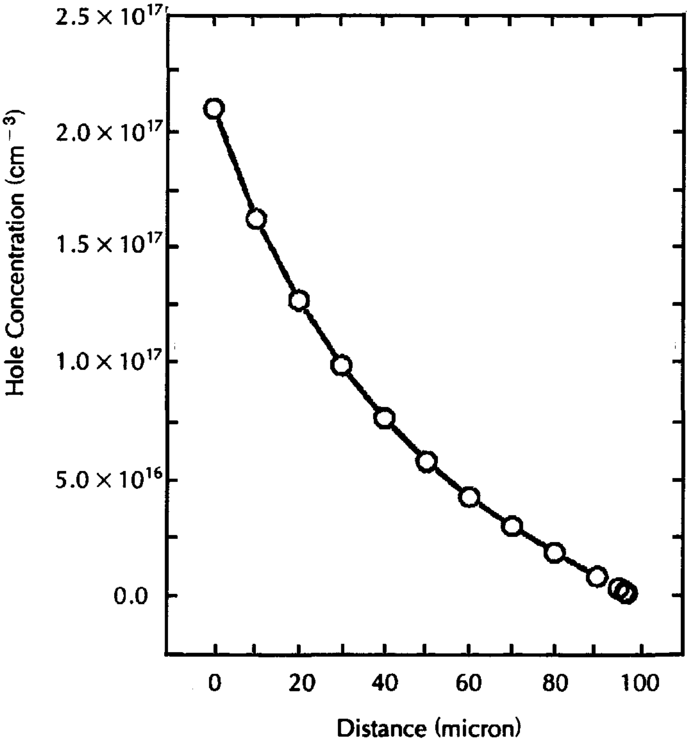


Figure E5.1.2 Hole concentration profile.

In the carrier distribution profile of a PIN rectifier in the ON state, the injected carrier concentration first falls as in IGBT but begins to rise again as one approaches the N layer. This happens because carrier injection occurs from both the P- and N-end regions of the PIN rectifier into the intrinsic layer. This feature distinguishes the IGBT from the PIN rectifier and accounts for the smaller forward voltage drop obtained with a PIN rectifier than the IGBT.

5.1.3 Forward Voltage Drop of IGBT

Utilizing the ON-state carrier distribution in the drift region of IGBT, the forward voltage drop of the IGBT is derived by summing the contributions from the junction between N^--drift region and P^+ substrate, the N^--drift region itself, and the MOSFET components (the channel, the JFET region, and the accumulation region). Here we shall consider the junction and N^--drift region components; the MOSFET components will be dealt with in Section 5.2.2.

Potential Drop Across the Forward-Biased P^+ Substrate/N^--Drift Region Junction. The equilibrium hole concentration in the N^--drift region is $p_{0\mathrm{N}^-} = n_\mathrm{i}^2/N_\mathrm{D}$. Then the voltage drop across the forward-biased junction between P^+ substrate and N^--drift region is expressed as [Eq. (4.6)]

$$V_{\mathrm{P^+N}} = \frac{kT}{q} \ln\left(\frac{p_0}{p_{0\mathrm{N}^-}}\right) \tag{5.12}$$

Using the p_0 equation (5.10), we have

$$V_{\mathrm{P^+N}} = \frac{kT}{q} \ln\left\{\frac{JL_\mathrm{a}N_\mathrm{D}}{2qD_\mathrm{p}n_\mathrm{i}^2} \tanh\left(\frac{d}{L_\mathrm{a}}\right)\right\} \tag{5.13}$$

Potential Drop Across the N^--Drift Region. Voltage drop across the N^--drift region of the IGBT is derived by integration of electric field distribution through the drift region. For obtaining the electric field distribution, the high-level injection condition $n(x) = p(x)$ is used along with the current transport equations (Appendices 4.1 and 4.2):

$$J_\mathrm{p} = q\mu_\mathrm{p}\left(pE - \frac{kT}{q}\frac{dp}{dx}\right) \tag{5.14}$$

and

$$J_\mathrm{n} = q\mu_\mathrm{n}\left(nE + \frac{kT}{q}\frac{dn}{dx}\right) \tag{5.15}$$

In the drift layer, the total current density is expressed as

$$J = J_n(x) + J_p(x) = \text{Constant} \tag{5.16}$$

Based on the above equations, the electric field distribution is expressed as

$$E(x) = \frac{J}{qp(\mu_n + \mu_p)} - \frac{kT}{q}\left(\frac{\mu_n - \mu_p}{\mu_n + \mu_p}\right)\left(\frac{1}{p}\right)\left(\frac{dp}{dx}\right) \tag{5.17}$$

Substituting for $p(x)$ from Eq. (5.11), we get

$$E(x) = \frac{kT}{qL_a}\left[\frac{2\mu_p}{(\mu_n + \mu_p)}\frac{\cosh(d/L_a)}{\sinh\{(d-x)/L_a\}} + \frac{\mu_n - \mu_p}{\mu_n + \mu_p}\frac{1}{\tanh\{(d-x)/L_a\}}\right] \tag{5.18}$$

By integrating this equation from $x = 0$ to $x = d$, the voltage drop in the N-drift region is

$$V_M = \frac{kT}{q}\left[\frac{2\mu_p}{\mu_n + \mu_p}\cosh(d/L_a)\ln\{\tanh(d/2L_a)\} + \frac{\mu_n - \mu_p}{\mu_n + \mu_p}\ln\{\sinh(d/L_a)\}\right] \tag{5.19}$$

Example 5.2 In an IGBT, resistivity of the N^--drift region is 30Ω-cm, its thickness is 50 μm and ambipolar diffusion length in this region is 30 μm. If P-base junction depth $X_{jP} = 3$ μm, calculate the total potential drop across the P^+N junction and the N^--drift region of IGBT operating at a forward voltage of 2 V and current density of 100 A/cm^2.

Figure E5.2.1 shows the IGBT unit cell. Since, $\rho = 30$ °-cm, $N_D = 1/(\rho e\mu) = 1/(30 \times 1.6 \times 10^{-19} \times 1350) = 1.54 \times 10^{14}$ cm^{-3}. For operation at 2 V, quasi-neutral

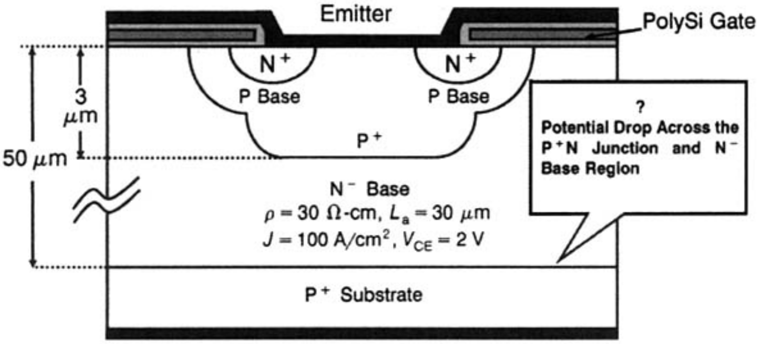

Emitter
PolySi Gate
N+
N+
P Base
P Base
P+
3 μm
50 μm
?
Potential Drop Across the P+N Junction and N- Base Region
N- Base
ρ = 30 Ω-cm, La = 30 μm
J = 100 A/cm², VCE = 2 V
P+ Substrate
Collector

Figure E5.2.1 IGBT for Example 5.2.

width of the N$^-$-drift region $= 50\times10^{-4} - 3\times10^{-4} - \sqrt{[\{2\varepsilon_0\varepsilon_s(V-\phi_0)\}/(qN_D)]} = 47\times10^{-4} - \sqrt{[\{2\times8.854\times10^{-14}\times11.9\times(2-0.7)\}/(1.6\times10^{-19}\times1.54\times10^{14})]} = 43.67\times10^{-4}$ cm. Now, voltage drop across the forward biased P$^+$N junction is $V_{P+N} = 0.0259\ln[\{(100\times30\times10^{-4}\times1.54\times10^{14})/(2\times1.6\times10^{-19}\times12.43)(1.45\times10^{14})2\}\tanh\{(43.67\times10^{-4})/(30\times10^{-4})\}] = 0.63781$ V. Voltage drop across the N$^-$-drift region is $V_M = 0.0259[\{(2\times480)/(1350+480)\}\cosh(43.67/30)\ln\{\tanh(43.67/60)\} + \{(1350-480)/(1350+480)\}\ln\{\sinh(43.67/30)\}] = -0.005895$ V. So the total voltage drop is $0.63781 - 0.005895 = 0.6319$ V.

5.1.4 Two-Dimensional Models of Carrier Distribution in the ON State

Carrier distribution in the N$^-$ base of the IGBT results directly from two-dimensional current flow [2]. Inspection of current flow in the IGBT half-cell of Fig. 5.3 shows that electron and hole currents, J_{ny} and J_{py}, entering the N$^-$ base at the P$^+$-substrate/N$^-$-base junction at $y = d_B$ flow homogeneously in the negative y-direction in both segments I and II of the half-cell. The reverse-biased P-base/N$^-$-base junction collects the holes so that the holes entering part IB at $y = d_B$ leave it at the P-base/N$^-$-base junction at $y = d_C$ but the holes entering part IIB and flowing into part IIA cannot penetrate the accumulation layer at the gate ($d_C \geq y \geq 0$), so they move laterally in the negative x direction. At the same time, under the influence of the electric field at the P-base/N$^-$-base junction, the electrons of part IB are compelled to flow laterally into part IIB and then into IIA, resulting in a high current density in this region. Because this elevated

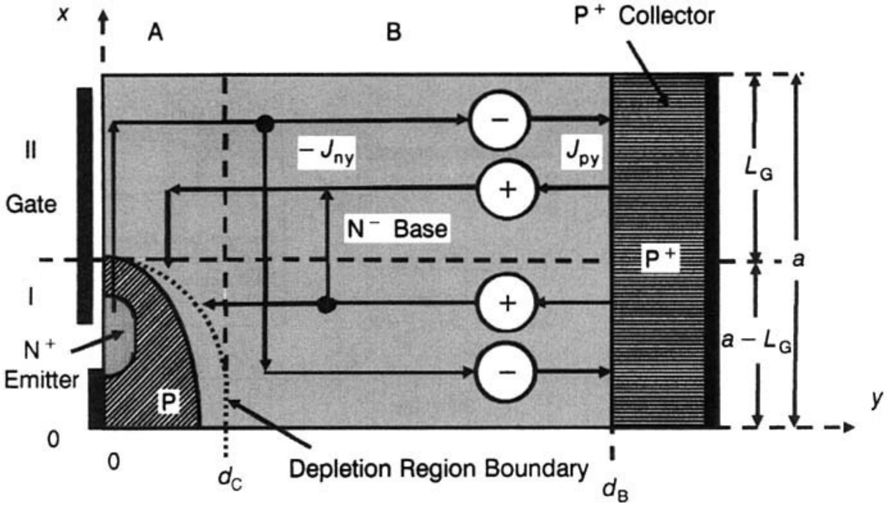


Figure 5.3 Two-dimensional current flow in the half-cell of a conventional IGBT.

electron concentration, and hence high current density region, has a smaller conductivity modulation in part II of the structure, a large voltage drop is produced here. This voltage drop is reduced by increasing the gate length L_G at a given half-cell width a. The outcome of this geometrical modification is that the current crowding is weakened in the region underneath the gate. As a result, the well-known "JFET effect," producing a significant fraction of the ON resistance, is decreased. Also, the conductivity modulation in this high-current-density region is enhanced. Both these factors act in conjunction to lower the conduction voltage drop.

To get a fundamental insight into the coupling of this electron and hole current flow in the vicinity of the gate, various models have been proposed based on two approaches. In the first approach, two one-dimensional models are developed which are combined together by lateral electron flow in front of the N^+ emitter and subsequently improved by considering lateral hole flow also. In the second approach, solution of the Laplace equation in the N base of IGBT, after simplifying the geometry by conformal mapping, leads to a two-dimensional (2-D) model. The 2-D model gives accurate dependence of carrier concentration in the proximity of the P base on L_G/a ratio with a realistic carrier density at the gate. The IGBT half-cell is transferred to the complex plane $z = x + jy$. By applying *Schwarz–Christoffel Transformation* [3, 4], the appropriate region in the z plane is mapped into the upper half of the image plane $w = u + jv$. A further such transformation maps the upper half of the w plane onto the region of the image plane $t = r + js$ shown in Fig. 5.4a. The interested reader may consult reference 2 to understand the mapping process. This arrangement is interpreted as a trench-IGBT with a cell width of 1, a base width s_B, and a gate length r_G. The carrier distribution in a point (suppose X) of the t plane comprises two parts. The first part is proportional to the sum of the solid angles Ω_i under which the gates of the periodically repeated trench cells appear in X. The second part varies linearly with the distance s between X and the real r axis where the gates and the emitters of the trench IGBT are situated. It has a positive or negative sign depending upon the carrier concentration in the N^- base. The obvious implication is that a horizontal carrier profile can be obtained in a major part of the base. It is also possible to achieve a profile with carrier concentration decreasing with increasing s values, which would give a higher conductivity modulation in regions where the current density is large. This will minimize the forward drop and thus improve the ON-state characteristics. For example, by increasing Ω_i, the conductivity modulation in N^- base is enhanced. Since Ω_i is proportional to r_G, a larger gate length is preferred. This allows adjustments of the carrier profile in the N^- base through the normalized length parameter L_G/a for conventional IGBT and r_G for the trench-IGBT. The trench IGBT structure displayed in Fig. 5.4b has an extremely short emitter length. Numerical simulations have shown that the carrier density near the gate is much higher than at the P^+ collector. Also, it is approximately constant in the vicinity of the gate and exhibits a pro-

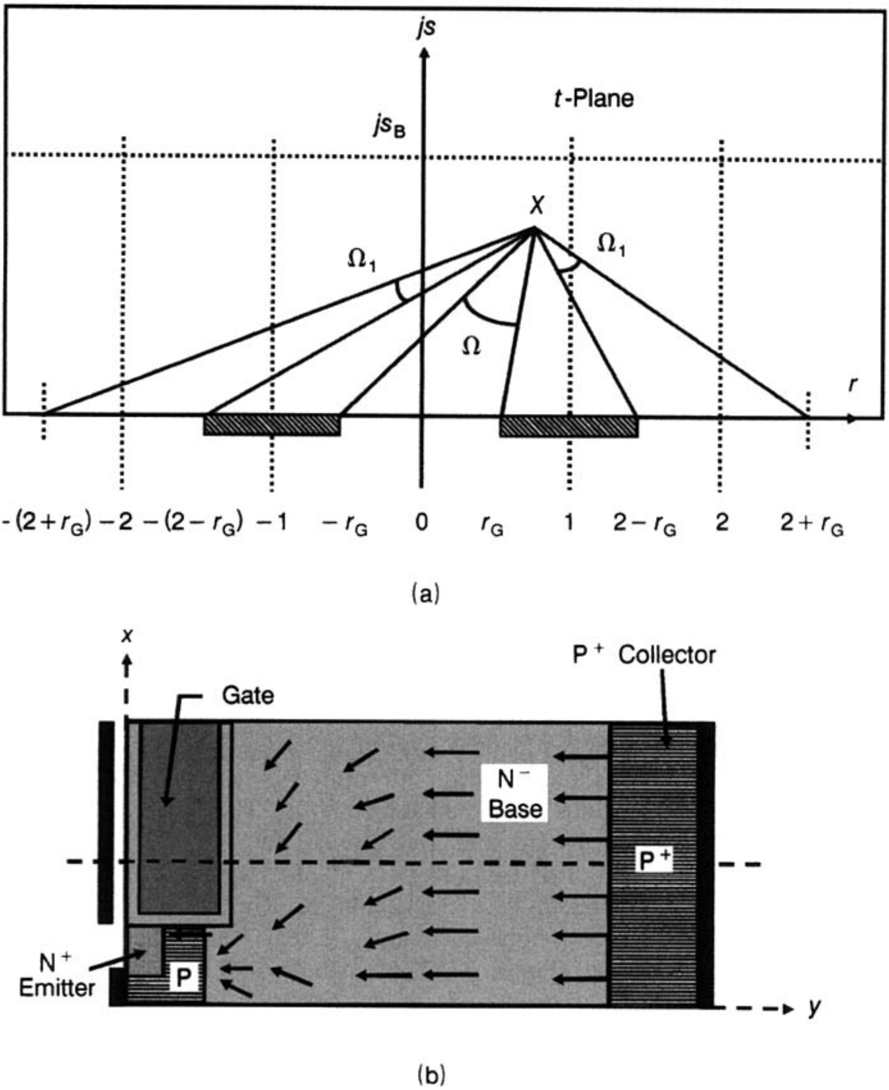

js
t-Plane
js_B
X
Ω_1
Ω
Ω_1
r
-(2+r_G) −2 −(2−r_G) −1 −r_G 0 r_G 1 2−r_G 2 2+r_G
(a)
x
Gate
P+ Collector
N− Base
P+
N+ Emitter
P
y
(b)

Figure 5.4 Analysis of carrier distribution in a point X of the complex t-plane. Mapped boundaries appear under the solid angles Ω_i. **Notation:** $t = r + js$, image plane; r_G, gate length; s_B, base width; s, distance between X and real axis; Ω_i: solid angles under which gates of trench cells appear in point X.

nounced fall near the emitter. Thus control of the h parameter h_E of the P^+ collector along with the gate length offers a convenient way of adjustment of the N^--base carrier profile and thereby static and dynamic properties of both types of IGBTs. ON-state voltage drop, of a conventional IGBT, with parameter values $a = 25\ \mu m$, $d_B = 250\ \mu m$, $d_C = 5\ \mu m$, and $L_G/a = 0.25$ is 2.12 V at 100 A/cm^2, but the same can be decreased to 1.58 V by using an

L_G/a ratio = 0.75 without altering the remaining device parameters [2]. The reason of the smaller potential drop is the readjustment of the carrier profile in the ON state, decreasing the JFET effect and increasing the conductivity modulation in high current density zone of the IGBT cell, as outlined previously.

If 2-D effects are neglected, a gross overestimation of the IGBT base voltage drop results leading to erroneous prediction of ON-state voltage. An analytical model for improved ON-state characteristics prediction has been proposed [5], taking into account the 2-D effects in forward conduction as well as temperature effects. The results of this model are in good agreement with those of 3-D simulations and experiments. The model is useful for IGBT structural optimization and circuit simulations.

5.2 BIPOLAR TRANSISTOR-DMOSFET MODEL OF IGBT BY EXTENSION OF PIN RECTIFIER-DMOSFET MODEL

5.2.1 Forward Conduction Characteristics

This model [1-6] views the IGBT as a MOSFET driving a wide-base PNP transistor in a Darlington configuration. Therefore the BJT equations are used in conjunction with MOSFET equations to describe the I–V characteristics. A close scrutiny of the IGBT structure (Fig.5.5a) keeping in mind its MOSFET and PNP transistor components reveals that the emitter of the PNP transistor is the collector of IGBT and collector of PNP transistor is the P base of IGBT. So, in the ensuing discussion, the emitter, base, and collector terminals must be related to the concerned transistor (PNP transistor or IGBT). In the PNP transistor, the collector represents the drain of the MOSFET. From this collector of the PNP transistor, electronic current supplied by MOSFET channel enters the N^- base of this transistor. So, the darkened boundary between collector and base of PNP transistor represents the connection between the drain of MOSFET and base of PNP transistor. During forward conduction of IGBT, this junction is reverse-biased, having zero excess carrier concentration. The dark boundary is called the *virtual base contact* of the PNP transistor (Fig. 5.5b). Electronic current entering the base of the PNP transistor = Base current of this transistor = MOSFET drain current, and Hole current entering the collector of PNP transistor = Collector current of this transistor. Therefore the electron current I_n flowing through the MOSFET channel of the IGBT (i.e., the base current of PNP transistor) is related to the hole current, I_p, flowing across the PNP transistor (collector current of this transistor) as [Eq. (4.64)]:

$$I_p = \beta_{PNP} I_n = \left(\frac{\alpha_{PNP}}{1 - \alpha_{PNP}} \right) I_n \tag{5.20}$$

where β_{PNP} denotes the common-emitter current gain of PNP transistor and

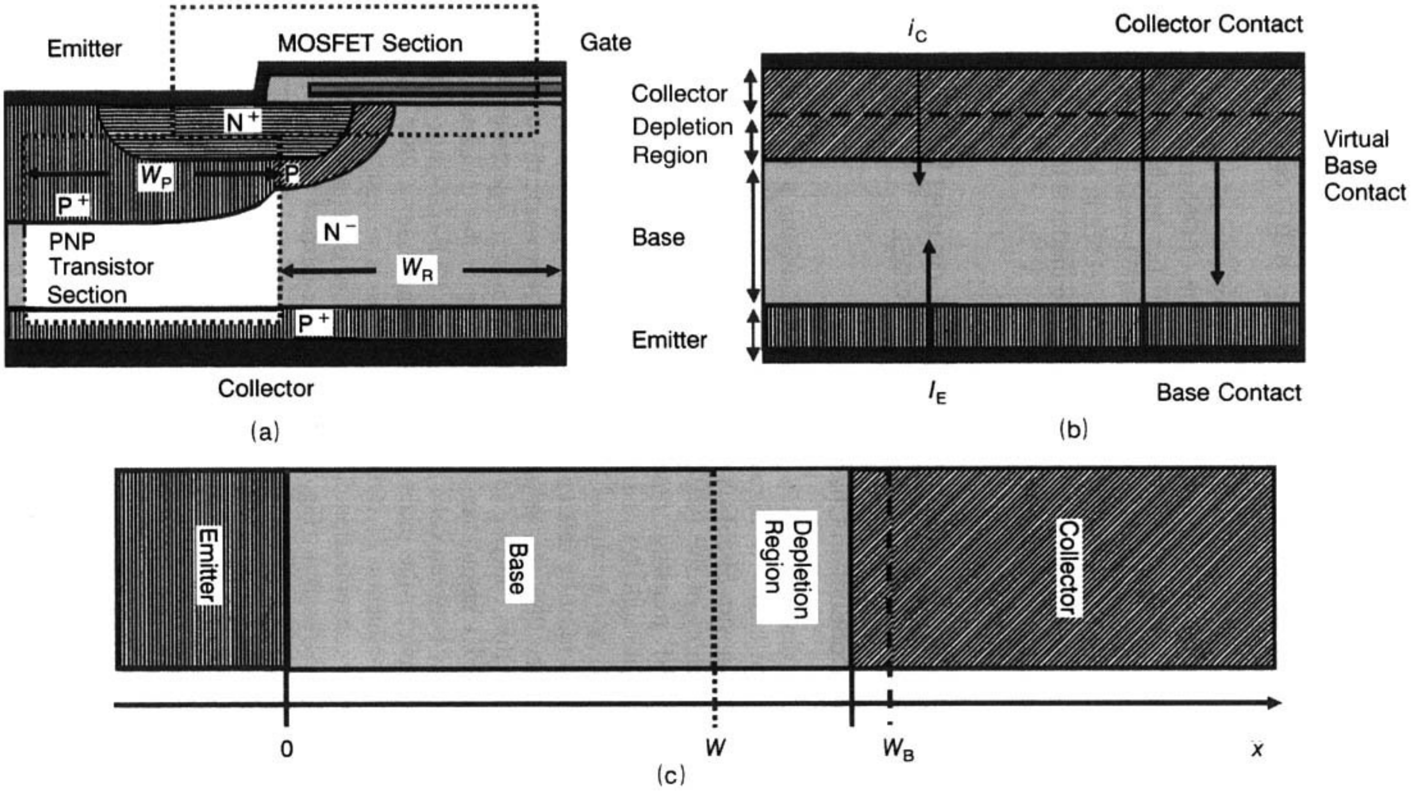

Emitter
MOSFET Section
Gate
N^+
W_P
P
P^+
PNP Transistor Section
N^-
W_R
P^+
Collector
(a)
i_C
Collector Contact
Collector
Depletion Region
Base
Emitter
Virtual Base Contact
I_E
Base Contact
(b)
Emitter
Base
Depletion Region
Collector
0
W
W_B
x
(c)

Figure 5.5 (a) Cross-sectional representation of IGBT showing the bipolar transistor and DMOSFET segments. (b) Carrier flow in the bipolar transistor of IGBT indicating the position of virtual base contact. (c) Coordinate system used for analysis.

equals $\alpha/(1-\alpha)$. The emitter current, I_E, is obtained by adding the electron and hole contributions as

$$I_E = I_n + I_p = \frac{I_n}{1-\alpha_{PNP}} \tag{5.21}$$

Due to the high input impedance of the MOS structure in an IGBT, gate current I_G is zero and the emitter current I_E approximately equals the collector current I_C. To derive the forward conduction characteristics of the IGBT, replacing I_{CE} by I_n modifies the forward drop Eq. (5.6) based on the PIN rectifier-DMOSFET model, yielding

$$V_{CE} = \left(\frac{2kT}{q}\right)\ln\frac{I_{CE}d}{2qW_R Z D_a n_i F(d/L_a)} + \frac{(1-\alpha_{PNP})I_{CE}L_{Ch}}{\mu_{ns}C_{ox}Z(V_{GE}-V_{Th})} \tag{5.22}$$

To examine how Eq. (5.22) is obtained, we note from Eq. (5.21) that $I_n = (1-\alpha_{PNP})I_E$. In Eq. (5.6), we put $(1-\alpha_{PNP})I_{CE}$ in place of I_{CE} in the second term because the collector current equals the emitter current due to the absence of gate current component in the high input impedance MOS structure. The overall impact is that the MOSFET channel current is a fraction of the total collector–emitter current. The drawback of the PIN rectifier–MOSFET model becomes evident here because it assumes that the full collector–emitter current flows through the MOSFET channel. But the above substitution of $(1-\alpha_{PNP})I_{CE}$ in place of I_{CE} is not necessary in the first term of Eq. (5.6) because the voltage drop across the N^--drift region is approximated by that across a PIN rectifier with a current I_{CE} flowing through it. A noteworthy observation is that the forward voltage drop across the IGBT, computed from this equation, is smaller than that obtained from the PIN rectifier–DMOSFET model of the IGBT. This is due to the collector current being subdivided between the MOSFET channel and P-base region, as pointed out.

Besides the forward potential drop, the saturated drain current $I_{CE,sat}$ of the IGBT and therefore the transconductance of IGBT, as calculated from the bipolar transistor–DMOSFET model of IGBT, is different from that given by the PIN rectifier–DMOSFET model. Upon decreasing the gate–emitter bias to such a value that the voltage drop in the channel limits the current flow, the electron current I_n can be written as [Eq. (3.26)]:

$$I_n = \frac{\mu_{ns}C_{ox}Z}{2L_{Ch}}(V_{GE}-V_{Th})^2 \tag{5.23}$$

Then from Eqs. (5.23) and (5.21), we obtain

$$I_E \approx I_{CE} = \frac{1}{1-\alpha_{PNP}}\frac{\mu_{ns}C_{ox}Z}{2L_{Ch}}(V_{GE}-V_{Th})^2$$

This is the saturated collector current so that

$$I_{\text{CE,sat}} = \frac{1}{1-\alpha_{\text{PNP}}} \frac{\mu_{\text{ns}} C_{\text{ox}} Z}{2L_{\text{Ch}}} (V_{\text{GE}} - V_{\text{Th}})^2 \tag{5.24}$$

and the transconductance in the saturation region is expressed as

$$\begin{aligned} g_{\text{ms}} &= \left.\frac{\partial I_{\text{CE,sat}}}{\partial V_{\text{GE}}}\right|_{V_{\text{DS}}=\text{Constant}} = \frac{1}{1-\alpha_{\text{PNP}}} \frac{\mu_{\text{ns}} C_{\text{ox}} Z}{2L_{\text{Ch}}} 2(V_{\text{GE}} - V_{\text{Th}}) \\ &= \frac{1}{1-\alpha_{\text{PNP}}} \frac{\mu_{\text{ns}} C_{\text{ox}} Z}{L_{\text{Ch}}} (V_{\text{GE}} - V_{\text{Th}}) \end{aligned} \tag{5.25}$$

It is obvious from this equation that the IGBT transconductance is comparatively higher than the MOSFET transconductance of same aspect ratio because of multiplication by the factor $1/(1-\alpha_{\text{PNP}})$. Since $\alpha_{\text{PNP}} \cong 0.5$, a ratio of 2 between the two transconductance values is common.

Example 5.3 Employing the first-order PIN rectifier–DMOSFET and bipolar transistor–MOSFET models of IGBT, calculate the ON-state voltage at current densities 0.1, 1, 10, 100, and 1000 A/cm^2 and plot the forward conduction characteristics (current density–voltage) of an IGBT in which width of N$^-$-drift region (d) is 50 μm, channel length (L_{Ch}) is 2 μm, and gate oxide thickness (t_{ox}) is 1000 Å at a gate bias of 10 V. The IGBT cell has a linear geometry. Its threshold voltage is 4 V. Distance between adjacent P-base wells (W_{R}) is 6 μm and total cell width ($W_{\text{R}} + W_{\text{P}}$) is 30 μm. It is given that $D_a = 18.34$ cm^2/sec. $F(d/L_a) = 0.25$, $\mu_{\text{ns}} = 500$ cm^2/V-sec, and $\alpha_{\text{PNP}} = 0.5$.

The schematic of IGBT unit cell is shown in Fig. E5.3.1. In 1 cm, the number of cells $= 1/(30 \times 10^{-4}) = 333.3$. The Z value is 333.3×1 cm $= 333.3$ cm for a cell of size 1 cm^2. Hence, $W_{\text{R}} = 333.3 \times 6 \times 10^{-4}$ cm. Also, $C_{\text{ox}} = 8.854 \times 10^{-14} \times 3.9/(1000 \times 10^{-8}) = 3.45306 \times 10^{-8}$ F/cm. Substituting the values in Eq. (5.6) for $J_{\text{CE}} = 0.1$ A, we have $V_{\text{F}} = 2 \times 0.0259 \ln\{(0.1 \times 50 \times 10^{-4})/(2 \times 1.6 \times 10^{-19} \times 6 \times 10^{-4} \times$

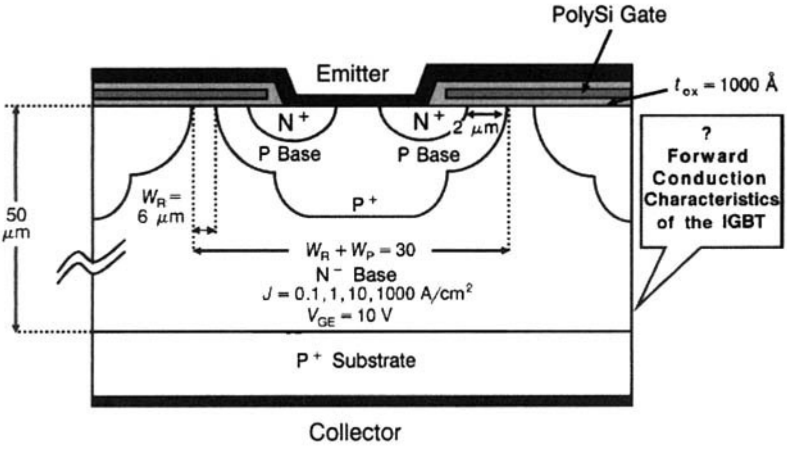


Figure E5.3.1 IGBT for Example 5.3.

Table E5.3.1 Forward Drops Across IGBT at Different Current Densities

		V_F(V)	
Serial No.	J_{CE} (A cm^{-2})	PIN Rectifier–DMOSFET Model	Bipolar Transistor–DMOSFET Model
1	0.1	0.3044	0.3041
2	1.0	0.429	0.426
3	10	0.6	0.57
4	100	1.24	0.95
5	1000	6.57	3.68

$333.3 \times 333.3 \times 18.34 \times 1.45 \times 10^{10} \times 0.25\} + (0.1 \times 2 \times 10^{-4})/(500 \times 3.45306 \times 10^{-8} \times 333.3 \times 6) = 0.3044$ V. In this manner, V_F is determined for $I_C = 1$, 10, 100, and 1000 A. Then the above calculations are repeated by applying Eq. (5.22). The values thus obtained are presented in Table E5.3.1.

The above data are plotted with V_F across the IGBT along the ordinate and J_{CE} along the abscissa. Then the X and Y axes are interchanged to convert the curves into current density-forward voltage ($J_{CE} - V_F$) characteristics, which are shown in Fig. E5.3.2.

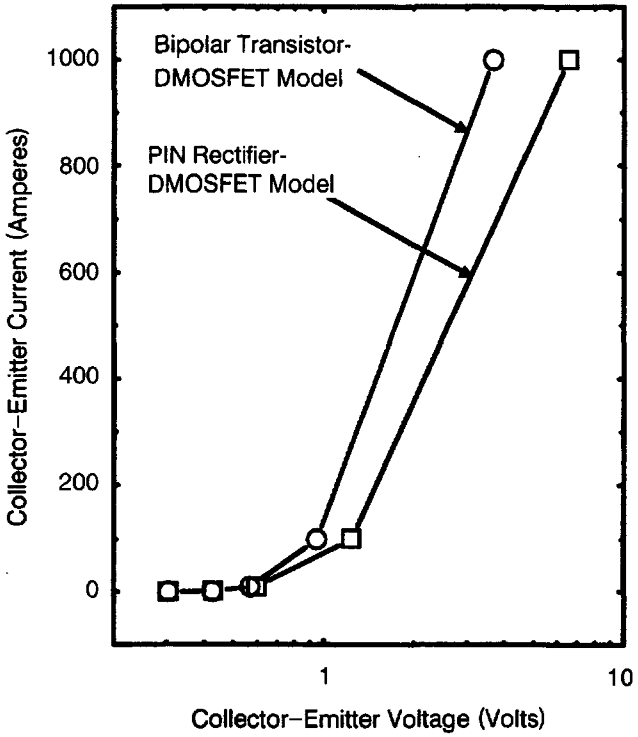


Figure E5.3.2 Comparison of *I–V* characteristics of IGBT calculated from the two models.

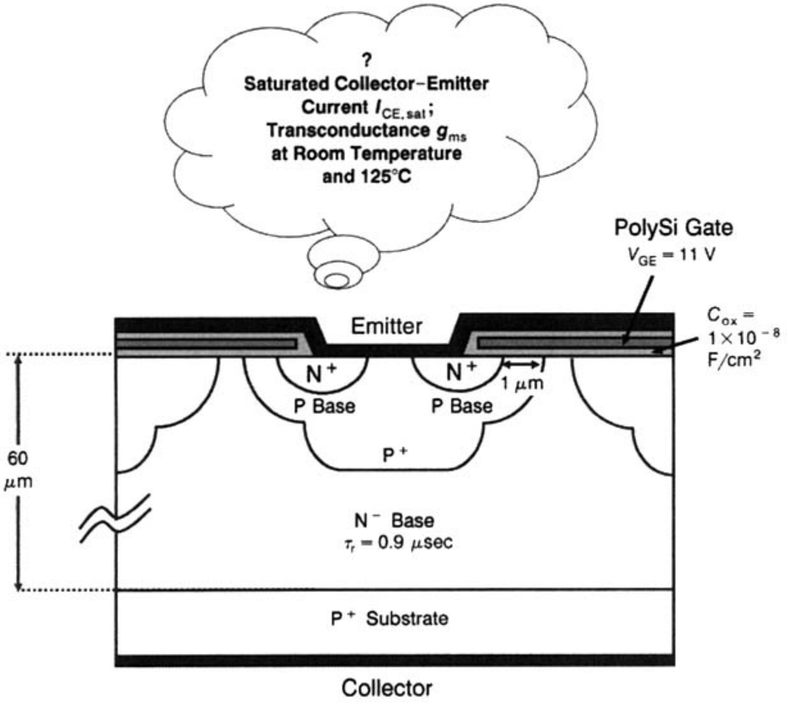

?
Saturated Collector-Emitter Current $I_{CE,sat}$; Transconductance g_{ms} at Room Temperature and 125°C
PolySi Gate
$V_{GE} = 11$ V
Emitter
$C_{ox} = 1\times10^{-8}$ F/cm^2
N+
N+
1 μm
P Base
P Base
P+
60 μm
N- Base
$\tau_r = 0.9$ μsec
P+ Substrate
Collector

Figure E5.4.1 IGBT for Example 5.4.

Example 5.4 (a) An IGBT with channel length of 1 μm, channel width of 150 cm, drift region thickness = 60 μm, carrier lifetime = 0.9 μsec, threshold voltage = 4.6 V, and oxide capacitance per unit area = 1×10^{-8} F/cm^2 is operating in the saturation region. Calculate the collector current and transconductance at a gate drive of 11 V, given that electron mobility in the inversion layer is 400 cm^2/V-sec and ambipolar diffusion coefficient is 18.34 cm^2/sec.

(b) If in part (a) above, the chip temperature is 125 °C, what is the effect on $r_{CE,sat}$ and g_{ms} for static conditions?

Figure E5.4.1 presents the cross-sectional diagram of IGBT unit cell.

For an assumed injection efficiency = 1, $\alpha_{PNP} = \alpha_T = 1/\cosh(W/L_a) = 1/\cosh\{W/\sqrt{(D_a\tau_a)}\} = 1/[\cosh\{(60\times10^{-4})/\sqrt{(18.34\times0.9\times10^{-6})}] = 0.4341$. Applying Eq. (5.24) we get $I_{CE,sat} = \{1/(1-0.4341)\}\times400\times1\times10^{-8}\times\{150/(2\times10^{-4})\}\times(11-4.6)^2 = 217.14$ A. By Eq. (5.25), $g_{ms} = \{1/(1-0.4341)\}\times400\times1\times10^{-8}\times\{150/(1\times10^{-4})\}\times(11-4.6) = 67.86\ \Omega^{-1}$.

(b) Thermal analysis is performed by applying the following temperature-dependent properties of Si (Table E5.4.1):

Using the above data, Eq. (5.24) yields $I_{CE,sat} = \{1/(1-0.4273)\}\times197.31\times1\times10^{-8}\times\{150/(2\times10^{-4})\}\times(11-3.718)^2 = 137.02$ A. By Eq. (5.25), $g_{ms} = \{1/(1-0.4273)\}\times197.31\times1\times10^{-8}\times\{150/(1\times10^{-4})\}\times(11-3.718) = 37.63\ \Omega^{-1}$.

Table E5.4.1 Variation of Si Properties with Temperature

Serial No.	Electrical Parameter	Formula Used	Application	Parameter Value at 125 °C
1	Electron mobility	$\mu_n(T_j) = 1500\,(300/T_j)^{2.5}$	$\mu_n(398) = 1500 \times (300/398)^{2.5}$	739.92 cm^2/V-sec
2	Hole mobility	$\mu_p(T_j) = 450\,(300/T_j)^{2.5}$	$\mu_p(398) = 450 \times (300/398)^{2.5}$	221.98 cm^2/V-sec
3	Electron mobility in the inversion layer	$\mu_{ns}(T_j) = 400\,(300/T_j)^{2.5}$	$\mu_{ns}(398) = 400 \times (300/398)^{2.5}$	197.31 cm^2/V-sec
4	Electron diffusion coefficient	$D_n\,(T_j) = \mu_n kT_j/q$	$D_n(398) = 739.92 \times 1.38 \times 10^{-23} \times 398/(1.6 \times 10^{-19})$	25.4 cm^2/sec
5	Hole diffusion coefficient	$D_p(T_j) = \mu_p kT_j/q$	$D_p(398) = 221.98 \times 1.38 \times 10^{-23} \times 398/(1.6 \times 10^{-19})$	7.62 cm^2/sec
6	Ambipolar diffusion coefficient	$D_a(T_j) = 2D_n(T_j)\,D_p(T_j)/\{D_n(T_j) + D_p(T_j)\}$	$D_a(398) = 2 \times 25.4 \times 7.62/(25.4 + 7.62)$	11.72 cm^2/sec
7	Ambipolar (high-level) lifetime	$\tau_{HL}(T_j) = \tau_{HL0}(T_j/T_0)^{\tau_{HL1}}$ where $\tau_{HL1} = 1.5$	$\tau_{HL}(398) = 0.9 \times 10^{-6} \times (398/300)^{1.5}$	1.375×10^{-6} sec
8	Threshold voltage	$V_T(T_j) = V_{T0} + V_{T1} \times (T - T_0)$ where $V_{T1} = -9$ mV/K	$V_T(398) = 4.6 + (-9 \times 10^{-3}) \times (398/300)$	3.718 V
9	Ambipolar diffusion length (L_a)	$L_a = \sqrt{(D_a\tau_a)}$	$L_a = \sqrt{(11.72 \times 1.375 \times 10^{-6})} = 4.015 \times 10^{-3}$	= 40.15 μm
10	W/L_a ratio	—	$W/L_a = 60/40.15$	1.4943
11	Current gain of PNP transistor	$\alpha_{PNP} = 1/\cosh(W/L_a)$	$1/\cosh(W/L_a) = 1/\cosh(1.4943)$	0.4273

5.2.2 MOSFET Components of Forward Voltage Drop of IGBT

Voltage drop across the MOSFET within the IGBT structure is simply the sum of voltage drops across the channel (V_{Ch}), the JEFET region (V_{JFET}), and the accumulation layer (V_{Acc}):

$$V_{MOSFET} = V_{Ch} + V_{JFET} + V_{ACC} \tag{5.26}$$

The channel voltage drop is

$$V_{Ch} = \frac{(1 - \alpha_{PNP})JL_{Ch}W_{Cell}}{\mu_{ns}C_{ox}(V_{GE} - V_{Th})} \tag{5.27}$$

The voltage drop across the JFET region is

$$V_{\mathrm{JFET}} = \frac{\rho_{\mathrm{JFET}}(1-\alpha_{\mathrm{PNP}})J(x_{\mathrm{p}}+W_0)W_{\mathrm{Cell}}}{L_{\mathrm{G}}-2x_{\mathrm{p}}-2W_0} \tag{5.28}$$

where ρ_{JFET} is the resistivity of the JFET region, L_{G} is the gate length, x_{P} is depth of P base, and W_0 is the zero-bias depletion region width. For IGBTs with low doping density of drift region, W_0 is large, leading to a high V_{JFET} value. Voltage drop across the accumulation layer is

$$V_{\mathrm{ACC}} = \frac{K(1-\alpha_{\mathrm{PNP}})J(L_{\mathrm{G}}-2x_{\mathrm{p}}-2W_0)W_{\mathrm{Cell}}}{2q\mu_{\mathrm{nA}}C_{\mathrm{ox}}V_{\mathrm{GE}}} \tag{5.29}$$

The reader may notice the resemblance of these equations with the corresponding equations for the MOSFET given in Section 3.4.1 [Eqs. (3.33), (3.35), and (3.34)].

Example 5.5 In an IGBT, the DMOSFET cell parameters are: half polySi gate length = 7 μm, half unit cell width = 15 μm, channel length = 0.96 μm, P-base depth = 2.2 μm, gate oxide thickness = 800 Å, threshold voltage = 4 V, $\alpha_{\mathrm{PNP}} = 0.55$, and $\rho_D = 30$ Ω-cm. If the zero-bias depletion layer is 3 μm, the gate bias is 15 V and the current density is 100 A/cm^2, compare the voltage drops across (i) the channel, (ii) the JFET, and (iii) accumulation regions. Carrier mobility in the inversion layer is 400 cm^2/V-sec and in the accumulation layer is 800 cm^2/V-sec.

The IGBT half-cell is shown in Fig. E5.5.1. Oxide capacitance per unit area is $8.854 \times 10^{-14} \times 11.9/(800 \times 10^{-8}) = 1.317 \times 10^{-7}$ F/cm^2. The required potential drops are determined below: (i) $V_{\mathrm{Ch}} = \{(1-0.55) \times 100 \times 0.96 \times 10^{-4} \times 30 \times 10^{-4}\}/$

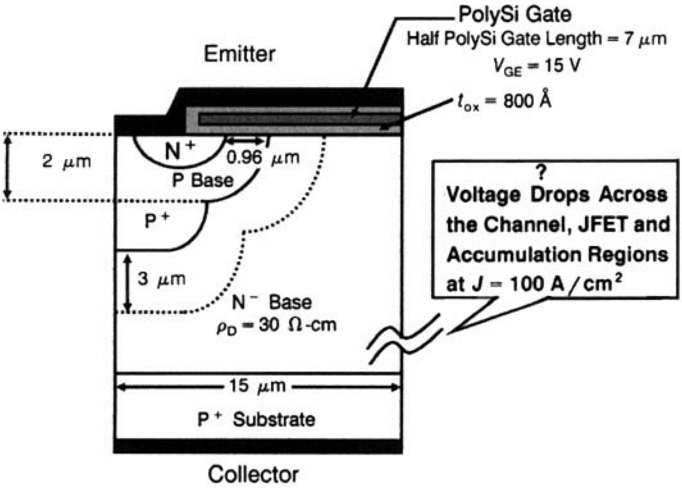

PolySi Gate
Half PolySi Gate Length = 7 μm
V_GE = 15 V
t_ox = 800 Å
Emitter
2 μm
N+
0.96 μm
P Base
P+
3 μm
N− Base
ρ_D = 30 Ω-cm
?
Voltage Drops Across the Channel, JFET and Accumulation Regions at J = 100 A/cm²
15 μm
P+ Substrate
Collector

Figure E5.5.1 IGBT for Example 5.5.

$\{400 \times 1.317 \times 10^{-7} \times (15 - 4)\} = 0.0224$ V. (ii) $V_{\mathrm{JFET}} = \{30 \times (1 - 0.55) \times 100 \times (2.2 + 3) \times 10^{-4} \times 30 \times 10^{-4}\}/\{(14 - 2 \times 2.2 - 2 \times 3) \times 10^{-4}\} = 5.85$ V. (iii) $V_{\mathrm{ACC}} = 0.6 \times (1 - 0.55) \times 100 \times (14 - 2 \times 2.2 - 2 \times 3) \times 10^{-4} \times 30 \times 10^{-4}\}/\{2 \times 800 \times 1.317 \times 10^{-7} \times (15 - 4)\} = 0.01258$ V.

5.2.3 Finite Collector Output Resistance of IGBT

It is expected that when the potential drop across the MOSFET channel is $> (V_{\mathrm{GE}} - V_{\mathrm{Th}})$, the collector–emitter current I_{CE} will saturate and become constant and invariant with V_{CE} changes. In practice, such saturation of I_{CE} is not observed. Instead, I_{CE} slightly increases with V_{CE} , at high values of V_{CE}. As a result, the collector output resistance does not become infinite but is restricted to a finite value. There are two reasons for divergence from this expected behavior, one due to the MOSFET component and the other due to the bipolar component of the IGBT. As V_{CE} increases, the MOSFET channel length decreases. This is accompanied by a decrease in channel resistance and rise of I_{CE} with V_{CE}. Furthermore, the rise of V_{CE} causes the depletion region of the PNP transistor collector–base junction to stretch whereby the undepleted base width of this transistor decreases, and hence the current gain, α_{PNP}, increases. The rise of current gain α_{PNP} is brought about by the decrease in undepleted base width W. The component of α_{PNP} responsible for this effect is the base transport factor d_{T} given by (Appendix 4.8)

$$\alpha_{\mathrm{T}} = 1 \Big/ \left\{ \cosh\left(\frac{W}{L_{\mathrm{a}}} \right) \right\} \tag{5.30}$$

where L_{a} is the ambipolar diffusion length in the N^{-}-drift region. At a given collector–emitter voltage V_{CE}, the undepleted base width is written as

$$W = d - \sqrt{\frac{2\varepsilon_0 \varepsilon_{\mathrm{s}} V_{\mathrm{CE}}}{qN_{\mathrm{D}}}} \tag{5.31}$$

where ε_0, ε_{s} are the permittivities of vacuum and Si, and N_{D} the doping concentration of N^{-}-drift region of IGBT. Eventually, this increase of current gain results in increase of I_{CE} with V_{CE}, imparting to the IGBT, a finite value of collector output resistance. At low V_{CE} values, W is large and its variation with V_{CE} is imperceptible. So in this portion of I_{CE}–V_{CE} characteristics, the effect of V_{CE} on W is insignificant. But at high V_{CE} values, W becomes small and changes rapidly with V_{CE}. Consequently, the influence of V_{CE} on W is more marked.

Ignoring the effect of channel length reduction and considering only the increase in current gain of the PNP transistor, the collector output resistance of IGBT is given by the equation

$$r_{\mathrm{C}}^{-1} = \frac{dI_{\mathrm{CE}}}{dV_{\mathrm{CE}}} = \frac{1}{2\sqrt{2}} \sqrt{\frac{\varepsilon_0 \varepsilon_{\mathrm{s}}}{qN_{\mathrm{D}} V_{\mathrm{CE}}}} \frac{\mu_{\mathrm{ns}} C_{\mathrm{ox}} Z (V_{\mathrm{GE}} - V_{\mathrm{Th}})^2 \sinh(W/L_{\mathrm{a}})}{L_{\mathrm{Ch}} L_{\mathrm{a}} \{\cosh(W/L_{\mathrm{a}}) - 1\}^2} \tag{5.32}$$

Differentiation of Eq. (5.24) and using Eq. (5.31) gives Eq. (5.32), as shown in Example 5.6 below.

Example 5.6 If in Example 5.3, dopant density of N^- base $= 1.54 \times 10^{14}$ cm^{-3}, P-base junction depth = 3 μm, and ambipolar diffusion length = 30 μm, calculate and plot the collector output resistance of the IGBT as a function of collector–emitter voltage V_{CE} for $V_{CE} = 20$, 50, 75, 100, 150, and 200 V. Ascertain the variation of current gain α_{PNP} with V_{CE}.

Taking the differential coefficient of Eq. (5.24) with respect to V_{CE}, we have

$$
\begin{aligned}
{r_{CE}}^{-1} &= \left[\{(d/dV_{CE})\alpha_{PNP}\}/(1-\alpha_{PNP})^2\right] \mu_{ns} C_{ox}\{Z/(2L_{Ch})\}(V_{GE}-V_{Th})^2 \\
&= -\left\{(1/L_a)\sinh(W/L_a)/\cosh(W/L_a)^2\right\} \times \left\{\cosh(W/L_a)^2\right\}/\left\{\cosh(W/L_a)-1\right\}^2 \times \\
&\quad -\tfrac{1}{2}\sqrt{\{(qN_D)/(2\varepsilon_0\varepsilon_s V_{CE})\}}\,\{2\varepsilon_0\varepsilon_s/(qN_D)\}\mu_{ns}C_{ox}\{Z/(2L_{Ch})\} \\
&\quad \times (V_{GE}-V_{Th})^2 = \frac{1}{2\sqrt{2}}\sqrt{\{(\varepsilon_0\varepsilon_s)/(qN_D V_{CE})\}}\,\mu_{ns}C_{ox}(Z/L_{Ch})(V_{GE}-V_{Th})^2 \\
&\quad \times \sinh(W/L_a)/\left[L_a\{\cosh(W/L_a)-1\}^2\right].
\end{aligned}
$$

Now, at $V_{CE} = 20V$, depletion layer depth is 20.847 μm, undepleted base width W is 34.153 μm, and $W/L_a = 34.153/30 = 1.138$. Applying Eq. (5.32), we have

$$
\begin{aligned}
r_c^{-1} &= \{1/(2\times 1.414)\}\sqrt{\{(8.854\times 10^{-14}\times 11.9)/(8\times 1.6\times 10^{-19}\times 1.54\times 10^{14}\times 20)\}} \\
&\quad \times 500 \times 3.45306 \times 10^{-8} \times 333.3 \times (20-4)^2 \times \sinh(1.138)/ \\
&\quad \left[2\times 10^{-4}\times 30\times 10^{-4}\times\{\cosh(1.138)-1\}^2\right] = 15.2228\ \Omega^{-1}.
\end{aligned}
$$

Therefore $r_c = 6.569 \times 10^{-2}$ Ω. Assuming an injection efficiency of unity, the current gain $= 1/\{\cosh(W/L_a)\} = 1/\{\cosh(1.138)\} = 0.581$. Exactly the same sequence of steps is followed to calculate the above parameters for the other V_{CE} values. The results of these calculations are presented in Table E5.6.1, and the dependence of collector output resistance on V_{CE} is depicted in Fig. E5.6.1.

Table E5.6.1 Current Gain of PNP Transistor and Collector Output Resistance of IGBT at Different Collector–Emitter Voltages

Serial No.	V_{CE} (volts)	Depletion Layer Width W_d (μm)	Undepleted Base Width $W = (50 - 3 - W_d)$ μm	W/L_a Ratio	$r_c(\Omega)$	α_{PNP}
1	20	12.85	34.15	1.138	6.569×10^{-2}	0.581
2	50	20.53	26.47	0.882	4.82×10^{-2}	0.7066
3	75	25.33	21.67	0.7223	3.236×10^{-2}	0.7859
4	100	29.14	17.86	0.595	2.09×10^{-2}	0.8457
5	150	35.73	11.27	0.376	6.433×10^{-3}	0.933
6	200	41.28	5.72	0.191	9.683×10^{-4}	0.982

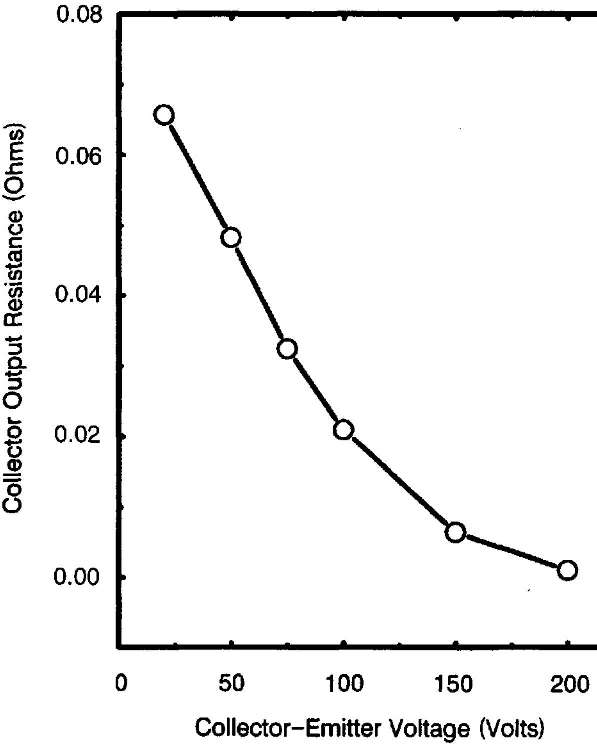


Figure E5.6.1 Dependence of collector output resistance of IGBT on collector-emitter voltage.

5.3 BIPOLAR TRANSISTOR-DMOSFET MODEL OF IGBT WITH DEVICE-CIRCUIT INTERACTIONS

In this model [6–10], both the steady-state and transient operation of the IGBT are analyzed, considering it as a circuit component. Analytical models, mixed mode device-circuit simulations, or network models for a circuit simulator are used for investigating the electrical behavior of the IGBT in a power circuit.

5.3.1 Steady-State Forward Conduction State

Ambipolar Carrier Transport Equations and Coordinate System for Analysis. The BJT in the IGBT is a lightly doped, wide-base, low-gain device operating under high-level injection for practical current densities. So ambipolar transport equations are used for describing carrier transport, for both steady-state and transient analysis. The electron current I_n and hole current I_h are expressed as (Appendix 5.2)

$$I_n = \frac{b}{1+b} I_{\text{CEI}} + qAD_a \frac{\partial p}{\partial x} \tag{5.33}$$

and

$$I_h = \frac{1}{1+b} I_{\text{CEI}} - qAD_a \frac{\partial p}{\partial x} \tag{5.34}$$

where b is the ambipolar mobility ratio $= \mu_n / \mu_p$, I_{CEI} is the total collector–emitter current, q is the electronic charge, A is the active area of the device, D is the ambipolar diffusivity, and p is the hole concentration. The time-dependent ambipolar diffusion equation is written as (Appendix 5.3)

$$\frac{\partial^2 \Delta p}{\partial x^2} = \frac{\Delta p}{L_a^2} + \frac{1}{D_a} \frac{\partial \Delta p}{\partial t} \tag{5.35}$$

where Δp is the excess hole concentration and L_a is the ambipolar diffusion length.

Figure 5.5c shows the rectangular Cartesian coordinate system used for analyzing the PNP transistor. The emitter of this transistor ends at $x = 0$. Its base extends from $x = 0$ to $x = d$, and undepleted or quasi-neutral base width at a given applied voltage stretches from $x = 0$ to $x = W$. So, $I_n(W)$ describes the PNP transistor base current and $I_p(W)$ describes its collector current.

Excess Carrier Concentration $\Delta p(x)$ and Excess Carrier Base Charge (Q_B) . During forward operation, the depletion-layer width W_d across the reverse-biased collector–base junction of the PNP transistor is given by

$$W_d = \sqrt{\frac{2\varepsilon_0 \varepsilon_s (V_a + \phi_0)}{qN_D}} \tag{5.36}$$

where V_a is the applied voltage, ϕ_0 is the built-in potential, and N_D is the N^--base doping concentration. This equation is obvious from Eq. (4.11). Then quasi-neutral base width is given by

$$W = d - W_d \tag{5.37}$$

Noting that $\Delta p(x)=0$ at $x=W$ and solving the steady-state ambipolar diffusion equation with $\partial \Delta p(x)/\partial t=0$ yields (Appendix 5.4)

$$\Delta p(x)=p_0 \sinh\left(\frac{W-x}{L_\mathrm{a}}\right)\bigg/\sinh\left(\frac{W}{L_\mathrm{a}}\right) \tag{5.38}$$

where $p_0=\Delta p(x)$ at $x=0$, that is, $p_0=$ excess carrier concentration at the base edge of the emitter–base junction of PNP transistor. Integration of excess carrier concentration through the base gives the total excess .carrier charge Q_B in the base:

$$\int_0^W \Delta p(x)\,dx=\frac{Q_\mathrm{B}}{qA}=\frac{p_0}{\sinh\left(\frac{W}{L_\mathrm{a}}\right)}\int_0^W \sinh\left(\frac{W-x}{L_\mathrm{a}}\right)dx$$

$$=\frac{p_0}{\sinh\left(\frac{W}{L_\mathrm{a}}\right)}\int_0^W\left\{\sinh\left(\frac{W}{L_\mathrm{a}}\right)\cosh\left(\frac{x}{L_\mathrm{a}}\right)-\cosh\left(\frac{W}{L_\mathrm{a}}\right)\sinh\left(\frac{x}{L_\mathrm{a}}\right)\right\}dx$$

$$=\frac{p_0}{\sinh\left(\frac{W}{L_\mathrm{a}}\right)}\left[\sinh\left(\frac{W}{L_\mathrm{a}}\right)\sinh\left(\frac{x}{L_\mathrm{a}}\right)\times\frac{1}{1/L_\mathrm{a}}\right.$$

$$\left.-\cosh\left(\frac{W}{L_\mathrm{a}}\right)\cosh\left(\frac{x}{L_\mathrm{a}}\right)\times\frac{1}{1/L_\mathrm{a}}\right]_{x=0}^{x=W}$$

$$=\frac{p_0L_\mathrm{a}}{\sinh\left(\frac{W}{L_\mathrm{a}}\right)}\left[-\cosh\left(\frac{W-x}{L_\mathrm{a}}\right)\right]_0^W=\frac{p_0L_\mathrm{a}}{\sinh\left(\frac{W}{L_\mathrm{a}}\right)}\left\{-1+\cosh\left(\frac{W}{l_\mathrm{a}}\right)\right\}$$

$$=\frac{p_0L_\mathrm{a}}{\sinh\left(\frac{W}{L_\mathrm{a}}\right)}\left\{\sinh^2\left(\frac{W}{2L_\mathrm{a}}\right)-\cosh^2\left(\frac{W}{2L_\mathrm{a}}\right)\right.$$

$$\left.+\sinh^2\left(\frac{W}{2L_\mathrm{a}}\right)+\cosh^2\left(\frac{W}{2L_\mathrm{a}}\right)\right\}$$

$$=\frac{p_0L_\mathrm{a}\times 2\sinh^2\left(\frac{W}{2L_\mathrm{a}}\right)}{2\sinh\left(\frac{W}{2L_\mathrm{a}}\right)\cosh\left(\frac{W}{2L_\mathrm{a}}\right)}$$

where the formulae $\sinh(A-B) = \sinh A \cosh B - \cosh A \sinh B$, $\cosh(A-B) = \cosh A \cosh B - \sinh A \sinh B$, $\int \sinh(au)\, du = (1/a)\cosh au$, $\sinh^2(A/2) + \cosh^2(A/2) = 1$, $\cosh A = \sinh^2(A/2) - \cosh^2(A/2)$, and $\sinh A = 2\sinh(A/2)\cosh(A/2)$ have been applied.

Therefore,

$$Q_B = qp_0 AL_a \tanh\left(\frac{W}{2L_a}\right) \tag{5.39}$$

This equation serves the initial condition for dynamic analysis.

Base and Collector Currents of the PNP Transistor. Based on the quasi-equilibrium approximation and under high-injection condition in the base, the electron current injected into the emitter is simply (Appendix 5.5)

$$I_n|_{x=0} = I_s\left(\frac{p_0^2}{n_i^2}\right) \tag{5.40}$$

where I_s denotes the emitter electron saturation current. Utilizing the ambipolar transport equations [Eqs. (5.33) and (5.34)] and the equations for $\Delta p(x)$ [Eq. (5.38)] and I_n $(x=0)$, [Eq. (5.40)], we obtain the equations for the base current $I_n(x=W)$ of PNP transistor and for the collector current $I_p(x=W)$, (Appendix 5.5):

$$I_B = I_s\left(\frac{p_0^2}{n_i^2}\right) + \left(\frac{qp_0 AD}{L_a}\right)\left\{\coth\left(\frac{W}{L_a}\right) - \frac{1}{\sinh(W/L_a)}\right\} \tag{5.41}$$

and

$$I_C = I_s\left(\frac{p_0^2}{bn_i^2}\right) + \left(\frac{qp_0 AD}{L_a}\right)\left\{\frac{\coth(W/L_a)}{b} + \frac{1}{\sinh(W/L_a)}\right\} \tag{5.42}$$

The total collector current of IGBT is $I_{CEI} = I_B + I_C$, which is an initial condition for transient analysis.

Emitter–Base Voltage Drop (V_{EB}) of the PNP Transistor. It is the sum of three components: (i) Potential drop across the emitter–base junction of this transistor at high-level injection, (ii) resistive drop across the conductivity-modulated base, and (iii) diffusion potential resulting from carrier distribution across the base under high-level injection, i.e.,

$$V_{\text{EB}} = \left(\frac{kT}{q}\right)\ln\left(\frac{p_0^{\,2}}{n_{\text{i}}^{\,2}}\right) + \frac{I_{CEI}W}{(1+1/b)\,\mu_{\text{n}}Aqn_{\text{eff}}} - \frac{D_{\text{a}}}{\mu_{\text{n}}}\ln\left(\frac{p_0+N_{\text{D}}}{N_{\text{D}}}\right) \tag{5.43}$$

where n_{eff} = effective low-doped base doping concentration given by (Appendix 5.6)

$$n_{\text{eff}} = \frac{W}{2L_{\text{a}}}\frac{\sqrt{N_{\text{D}}^{\,2}+p_0^{\,2}\text{cosech}^2(W/L_{\text{a}})}}{\tanh^{-1}\left[\dfrac{\sqrt{\{N_{\text{D}}^{\,2}+p_0^{\,2}\text{cosech}^2(W/L_{\text{a}})\}}\,\tanh(W/2L_{\text{a}})}{\{N_{\text{D}}+p_0\,\text{cosec}(W/L_{\text{a}})\}\tanh(W/2L_{\text{a}})}\right]} \tag{5.44}$$

Example 5.7 For the IGBT described in Example 5.2, find the emitter–base voltage V_{EB} by applying Eq. (5.43).

By Eq. (5.10), $p_0 = \{(100 \times 30 \times 10^{-4})/(2 \times 1.6 \times 10^{-19} \times 1.43)\}$ $\{\sinh(43.67/30)\}/\{\cosh(4367/30)\} = 6.764 \times 10^{16}\ \text{cm}^{-3}$. Now, the first term of Eq. (5.43) is 0.0259 $\ln\{(6.764 \times 10^{16})^2/(1.45 \times 10^{10})^2\} = 0.7954$ V. The third term is $(18.34/1350)\ln\{(6.764 \times 10^{16} + 1.54 \times 10^{14})/(1.54 \times 10^{14})\} = 0.082697$ V. Evaluation of the second term requires the determination of n_{eff}. We note that $W/L_{\text{a}} = 43.67/30 = 1.456$ and $W/(2L_{\text{a}}) = 43.67/60 = 0.7278$. So, $n_{\text{eff}} = [0.7278$ $\sqrt{\{(1.54 \times 10^{14})^2 + (6.764 \times 10^{16})^2 \times \text{cosech}^2(1.456)\}}\ ]/\text{arctanh}[\sqrt{\{(1.54 \times 10^{14})^2 + (6.764 \times 10^{16})^2 \times \text{cosech}^2(1.456)\}} \times \tanh\ (0.7278)$ $\{(1.54 \times 10^{14}) + (6.764 \times 10^{16})\tanh((0.7278)\}] = 8.42 \times 10^{15}\ \text{cm}^{-3}$. So, the second term $= (100 \times 43.67 \times 10^{-4})/(1.33 \times 1350 \times 1.6 \times 10^{-19} \times 8.42 \times 10^{15}) = 0.18054$ V. Thus the total voltage drop $V_{\text{EB}} = 0.7954 + 0.18054 - 0.082697 = 0.8932$ V.

Drain–Source Voltage Drop (V_{DS}) Across the MOSFET and the Forward Voltage Drop (V_{CEI}) of the IGBT. (V_{CEI}) = Emitter–base voltage drop (V_{EB}) of the PNP transistor + Collector–base voltage drop across this transistor (V_{CB}), or

$$V_{\text{CEI}} = V_{\text{EB}} + V_{\text{BC}} \tag{5.45}$$

The voltage V_{CB} is the drain–source voltage (V_{DS}) across the MOSFET. Because $I_{\text{B}} = I_{\text{MOSFET}}$ and the MOSFET is operating in the linear region [Eq. (3.22)], we obtain

$$V_{\text{CB}} = \frac{I_{\text{B}}}{K(V_{\text{GS}} - V_{\text{Th}})} \tag{5.46}$$

where

$$K = C_{\mathrm{ox}}\,\mu_{\mathrm{ns}}\left(\frac{Z}{L_{\mathrm{Ch}}}\right) \tag{5.47}$$

C_{ox} is the oxide capacitance per unit area, μ_{ns} is the surface electron mobility, Z is the channel width, and L_{Ch} is the channel length. Thus the IGBT ON-state voltage (V_{CEI}) is explicitly expressed in terms of the collector–emitter current $I_{\mathrm{CEI}} = I_{\mathrm{B}} + I_{\mathrm{C}}$ and gate voltage $V_{\mathrm{GEI}} = V_{\mathrm{GS}}$, assuming $V_{\mathrm{CB}} = 0$. The equation for V_{CEI} is another initial condition for studying the switching transient.

5.3.2 Dynamic Model of an IGBT and Its Switching Behavior

Reducing the gate–emitter voltage (V_{GEI}) below the threshold level turns off the IGBT. Then the MOSFET channel disappears and the base current I_{B} supplied to PNP transistor is stopped, that is, I_{B} becomes zero. The collector current of the PNP transistor I_{C} starts falling through carrier decay in the open-base bipolar transistor. Both the excess carrier base charge Q and the IGBT collector voltage V_{CI} are time-varying quantities. In the sequence, time-varying quantities will be marked by a prime to distinguish them from steady-state parameters.

Figure 5.6 depicts the main capacitances associated with the IGBT. The symbols are explained below. For the PNP transistor, C_{CER} is the collector–emitter redistribution capacitance:

$$C_{\mathrm{CER}} = \frac{W^2}{W_{\mathrm{eff}}{}^2}\,\frac{C_{\mathrm{BCJ}}Q}{3qAWN_{\mathrm{sc}}} \tag{5.48}$$

W_{eff} is the effective width for base transport and N_{sc} is the collector–base space-charge concentration. C_{EBJ} and C_{EBD} are the emitter–base junction and diffusion capacitances, and R_{B} is the conductivity-modulated base resistance of PNP transistor given by

$$R_{\mathrm{B}} = \frac{W}{(1 + 1/b)\,\mu_{\mathrm{n}} Aqn_{\mathrm{eff}}} \tag{5.49}$$

For the VDMOSFET, C_{GDJ} is the gate–drain overlap depletion capacitance:

$$C_{\mathrm{GDJ}} = \frac{\varepsilon_0\,\varepsilon_{\mathrm{s}}\,A_{\mathrm{GD}}}{W_{\mathrm{GDJ}}} \tag{5.50}$$

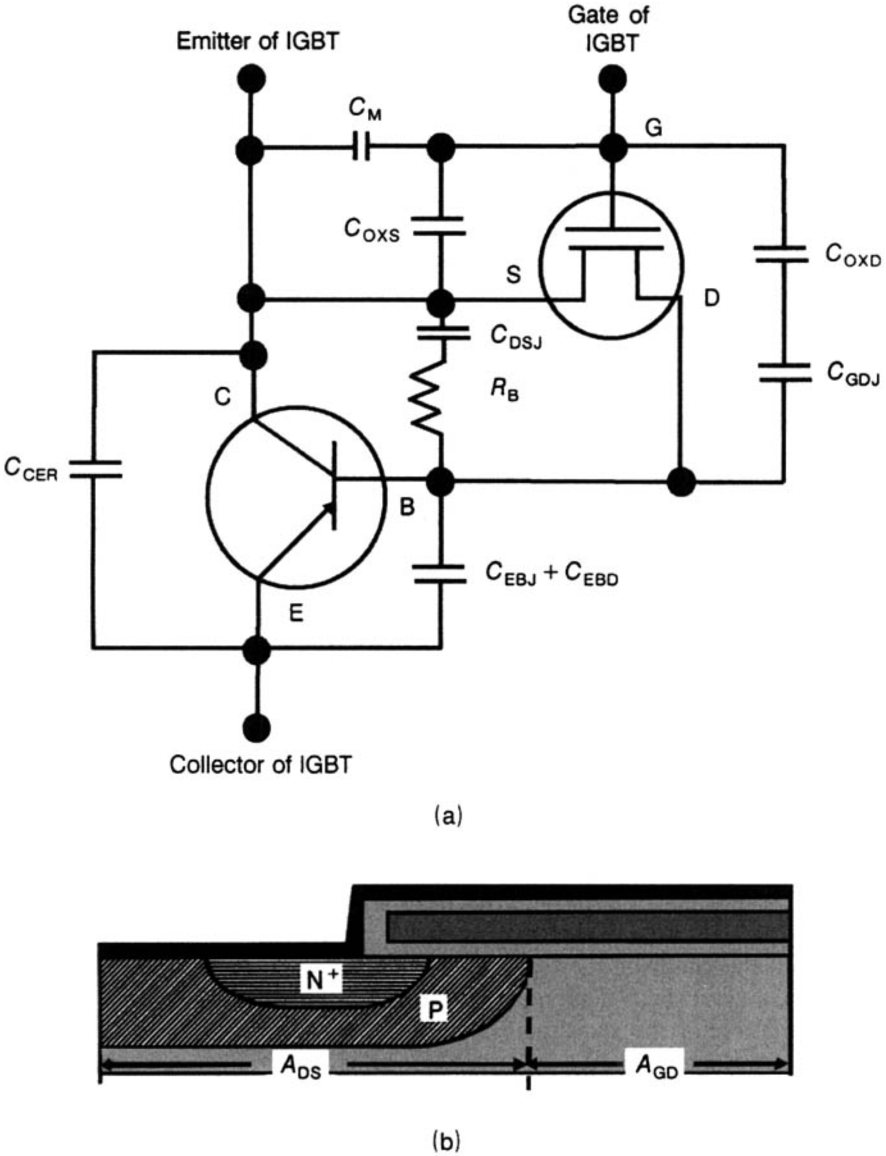

Emitter of IGBT
Gate of IGBT
C_M
G
C_{OXS}
S
D
C_{OXD}
C_{DSJ}
C_{GDJ}
R_B
C
C_{CER}
B
$C_{EBJ}+C_{EBD}$
E
Collector of IGBT
(a)
N+
P
A_{DS}
A_{GD}
(b)

Figure 5.6 (a) Bipolar and MOSFET components of the dynamic IGBT model. (b) Portion of IGBT cross section showing the areas covered under A_{DS} and A_{GD}.

C_{DSJ} is the drain–source overlap depletion capacitance:

$$C_{DSJ} = \frac{\varepsilon_0 \varepsilon_s (A - A_{GD})}{W_{DSJ}} \tag{5.51}$$

C_{OXS} is the gate oxide capacitance of the source overlap and C_{M} is the source metallization capacitance. C_{OXS} and C_{M} are the main components of the gate–source capacitance C_{GS}. The capacitance C_{OXD} is the gate oxide capacitance of the drain overlap, a dominant component of gate–drain capacitance C_{GD}.

A noteworthy observation is that during transient conditions, a quasi-static (QS) approximation is not valid for the IGBT. This is so despite the fact that rate of base charge variation is not very fast in PNP transistor. The violation of quasi-static approximation originates from the difference in boundary conditions on the electron and hole currents during the transient state from those in the steady state. As a result, the shape of the excess carrier distribution and the relation between current and total excess carrier base charge differ between these states. This contradicts the assumption in the quasi-static approximation, that the relation between current and total excess carrier charge in the base remains the same during transient and steady states. The dissimilarity arises due to the coupling of the electron and hole transport for the *low-gain, high injection conditions* of the total collector current I_{C}, such that neither can be treated independently. Under these conditions, the electron and hole concentrations are nearly equal in magnitude. Interaction occurs between the electrons flowing from the collector side of the base of PNP transistor toward the emitter and holes moving from the emitter toward the collector side of the base. Owing to the coupling effect, the collected hole current changes upon withdrawing from the MOSFET electron current. This kind of coupling is, however, debilitated for *high gain* or *low-level injection cases*. Furthermore, under common loading situations, during transitions of the collector–emitter voltage of IGBT, the depletion layer width across the collector–base junction of PNP transistor changes at a faster rate than the transit time of excess carriers through the base. So, the rate of change of the quasi-neutral base width is more rapid than the base transit speed, requiring a large proportion of current, for carrier redistribution, into the changing base width.

The collector–base depletion width decays with time, resulting in a transitory variation in the quasi-neutral base width. The rate of change of the quasi-neutral base width with time is expressed in terms of time rate of change of collector-base voltage as

$$\frac{dW}{dt} = -\frac{C_{\mathrm{D}}}{qN_{\mathrm{D}}A}\frac{dV_{\mathrm{BC}}}{dt} \tag{5.52}$$

where $C_{\mathrm{D}} = \varepsilon_0 \varepsilon_{\mathrm{s}} A/W_{\mathrm{d}}$ is the instantaneous collector–base depletion capacitance. Consequently, the excess carrier base charge Q is driven into a quasi-neutral base region, which shrinks with increase of voltage. This implies that the ambipolar diffusion equation must be solved for the moving quasi-neutral base width boundary condition resulting in the equation

$$\Delta p'(x) = p'_0(x)\left(1 - \frac{x}{W'}\right) - \left(\frac{p'_0}{W'D_a}\right)\left(\frac{x^2}{2} - \frac{W'x}{6} - \frac{x^3}{3W'}\right)\frac{dW'}{dt} \tag{5.53}$$

Integration of this equation through the base gives the total base charge under transient conditions as $Q' = qp_0AW'/2$ and gives the transient collector current at $x = W'$ as

$$I'_p(W') = \left(\frac{1}{1+b}\right)I'_{CEI} + \left(\frac{b}{1+b}\right)\frac{4D_pQ'}{W'^2} - \frac{Q'}{3W'}\frac{dW'}{dt} \tag{5.54}$$

The physical explanation of the three terms in this equation is given below: The first term is a non-quasi-static term, representing the variation of collector current of the IGBT with the instantaneous total current due to electron and hole transport coupling. Because the second term is related to the residual base charge and the applied collector–base voltage, it is the charge-controlled component of the current. The third term causes carrier redistribution for the moving boundary neutral base width boundary condition.

5.3.3 State Equations for IGBT Turn-Off Transient

For zero MOSFET current, we obtain

$$I_n\big|_{x=W} = C_{BCJ}\frac{dV_{BC}}{dt} \tag{5.55}$$

Since the total current at $x = W$ is the sum of electron and hole currents, we have

$$C'_{BCJ}\left(\frac{dV'_{BC}}{dt}\right) = I'_{CEI} - I'_P(W') \tag{5.56}$$

Furthermore, because the excess carrier base charge decays by recombination in the base and by electron current injection into the emitter, we can write

$$\frac{dQ'}{dt} = -\frac{Q'}{\tau_{HL}} - I_S\left(\frac{p'^2_0}{n_i^2}\right) \tag{5.57}$$

These equations are used to obtain dV_{BC}/dt and dQ_B/dt in terms of the instantaneous values of I_{CEI}, Q and V_{BC}. The analysis yields the non-quasi-static voltage state and charge state equations as follows:

$$\frac{dV'_{\mathrm{BC}}}{dt}=\frac{I'_{\mathrm{CEI}}-\dfrac{4D_{\mathrm{p}}Q'}{W'^2}}{C'_{\mathrm{BCJ}}(1+1/b)\left(1+\dfrac{Q'}{3qAN_{\mathrm{D}}W'}\right)} \tag{5.58}$$

and

$$\frac{dQ'}{dt}=-\frac{Q'}{\tau_{\mathrm{HL}}}-\frac{4Q'I_{\mathrm{s}}}{W'^2A^2q^2n_{\mathrm{i}}^2} \tag{5.59}$$

Note 1

For deriving Eq. (5.59), integration of Eq. (5.53) is performed through the base from $x=0$ to $x=W$, remembering that for a constant collector voltage the second term in Eq. (5.53) is zero. This gives p'_0 in terms of the total excess carrier base charge Q';

$$\int_0^{W'}\Delta p'\,dx=\frac{Q'}{qA}=p'_0\int_0^{W'}dx-\frac{p'_0}{W'}\int_0^{W'}x\,dx=p'_0W'-\frac{p'_0}{W'}\left[\frac{x^2}{2}\right]_0^{W'}$$

$$=p'_0W'-\frac{p'_0}{W'}\times\frac{W'^2}{2}=\frac{1}{2}p'_0W'$$

Hence,

$$p'_0=\frac{2Q'}{qAW'}$$

Substituting for p'_0 in Eq. (5.57), we get Eq. (5.59)

$$\frac{dQ'}{dt}=-\frac{Q'}{\tau_{\mathrm{HL}}}-I_{\mathrm{s}}\left(\frac{2Q}{qAW'}\right)^2\left(\frac{1}{n_{\mathrm{i}}^2}\right)=-\frac{Q'}{\tau_{\mathrm{HL}}}-\frac{4Q'I_{\mathrm{s}}}{W'^2A^2q^2n_{\mathrm{i}}^2}$$

Note 2

To obtain Eq. (5.58), we begin by differentiating Eq. (5.53), giving

$$\frac{d}{dx}\{\Delta p'(x)\}=p'_0\left(0-\frac{1}{W'}\right)-\left(\frac{p'_0}{W'D_{\mathrm{a}}}\right)\left(\frac{2x}{2}-\frac{W'}{6}-\frac{3x^2}{3W'}\right)\frac{dW'}{dt}$$

At $x = W'$,

$$
\begin{aligned}
\frac{d}{dx}\{\Delta p'(x)\}\bigg|_{x=W'} &= -\frac{p'_0}{W'} - \left(\frac{p'_0}{W'D_\mathrm{a}}\right)\left(\frac{2W'}{2} - \frac{W'}{6} - \frac{3W'^2}{3W'}\right)\frac{dW'}{dt} \\
&= -\frac{p'_0}{W'} - \left(\frac{p'_0}{W'D_\mathrm{a}}\right)\left(W' - \frac{W'}{6} - W'\right)\frac{dW'}{dt} \\
&= -\frac{p'_0}{W'} + \left(\frac{p'_0}{6D_\mathrm{a}}\right)\frac{dW'}{dt} = -\frac{\dfrac{2Q'}{qAW'}}{W'} + \left(\frac{\dfrac{2Q'}{qAW'}}{6D_\mathrm{a}}\right)\frac{dW'}{dt} \\
&= -\frac{2Q'}{qAW'^2} + \frac{Q'}{3qAD_\mathrm{a}W'}\frac{dW'}{dt}
\end{aligned}
$$

Substituting for $(d/dx)\{\Delta p'(x)\}$ into Eq. (5.34), we obtain

$$
\begin{aligned}
I_\mathrm{h} &= \frac{1}{1+b}I_\mathrm{CEI} - qAD_\mathrm{a} \times \left(-\frac{2Q'}{qAW'^2} + \frac{Q'}{3qAD_\mathrm{a}W'}\frac{dW'}{dt}\right) \\
&= \frac{1}{1+b}I_\mathrm{CEI} + 2\frac{D_\mathrm{a}Q'}{W'^2} - \frac{Q'}{3W'} \cdot \frac{dW'}{dt}
\end{aligned}
$$

The ambipolar diffusivity is

$$
D_\mathrm{a} = \frac{2D_\mathrm{n}D_\mathrm{p}}{D_\mathrm{n} + D_\mathrm{p}}
$$

from which

$$
\frac{D_\mathrm{a}}{D_\mathrm{p}} = \frac{2D_\mathrm{n}}{D_\mathrm{n} + D_\mathrm{p}} = \frac{2}{1 + \dfrac{D_\mathrm{p}}{D_\mathrm{n}}} = 2\frac{1}{1 + \dfrac{1}{b}} = 2\frac{b}{1+b} \quad \text{or} \quad D_\mathrm{a} = 2\frac{b}{1+b}D_\mathrm{p}
$$

Substituting for D_a into the equation for I_h and noting that $I_\mathrm{h} = I'_\mathrm{p}(W')$, we arrive at Eq. (5.54):

$$
\begin{aligned}
I'_\mathrm{p}(W') = I_\mathrm{h} &= \frac{1}{1+b}I_\mathrm{CEI} + 2\frac{2\dfrac{b}{1+b}Q'}{W'^2} - \frac{Q'}{3W'} \cdot \frac{dW'}{dt} \\
&= \frac{1}{1+b}I_\mathrm{CEI} + \left(\frac{b}{1+b}\right)\frac{4D_\mathrm{p}Q'}{W'^2} - \frac{Q'}{3W'} \cdot \frac{dW'}{dt}
\end{aligned}
$$

Substituting for $I'_p(W')$ into Eq. (5.56), we have

$$C'_{BCJ}\left(\frac{dV'_{BC}}{dt}\right) = I'_{CEI} - \frac{1}{1+b}I_{CEI} - \left(\frac{b}{1+b}\right)\frac{4D_pQ'}{W'^2} + \frac{Q'}{3W'}\cdot\frac{dW'}{dt}$$

$$= \frac{b}{1+b}I'_{CEI} - \left(\frac{b}{1+b}\right)\frac{4D_pQ'}{W'^2} + \frac{Q'}{3W'}\cdot\frac{dW'}{dt}$$

Employing Eq. (5.52), dW'/dt can be expressed in terms of dV_{BC}/dt and putting $C_D = C'_{BCJ}$,

$$C'_{BCJ}\left(\frac{dV'_{BC}}{dt}\right) + \frac{b}{1+b}I'_{CEI} - \left(\frac{b}{1+b}\right)\frac{4D_pQ'}{W'^2} - \frac{Q'}{3W'}\cdot\frac{C_{BCJ}}{qN_DA}\cdot\frac{dV'_{BC}}{dt}$$

or

$$C'_{BCJ}\left(\frac{dV'_{BC}}{dt}\right) + \frac{Q'}{3W'}\cdot\frac{C'_{BCJ}}{qN_DA}\cdot\frac{dV'_{BC}}{dt} = \frac{b}{1+b}I'_{CEI} - \left(\frac{b}{1+b}\right)\frac{4D_pQ'}{W'^2}$$

$$= \frac{1}{1+\dfrac{1}{b}}\left\{I'_{CEI} - \frac{4D_pQ'}{W'^2}\right\}$$

dV'_{BC}/dt is then

$$\frac{dV'_{BC}}{dt} = \frac{I'_{CEI} - \dfrac{4D_pQ'}{W'^2}}{C'_{BCJ}\left(1+\dfrac{1}{b}\right)\left\{1+\dfrac{Q'}{3qAN_DW'}\right\}}$$

which is Eq. (5.58).

Figure 5.7 shows the schematic of the circuit for series resistor–inductor load switching analysis. In this circuit, I_{CEI} is the total collector–emitter current, R is the series load resistance and LL the series load inductance,

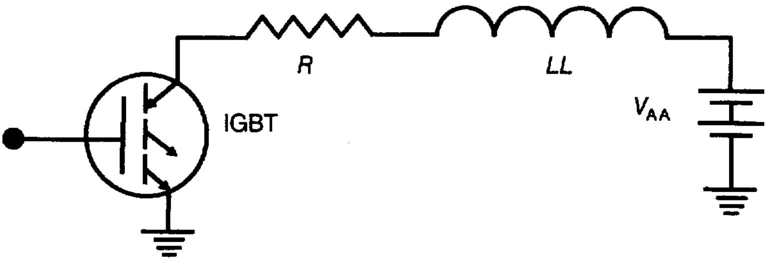


Figure 5.7 Circuit configuration for series resistor inductor load switching.

and V_{AA} is the collector supply voltage. The state equation for this circuit is

$$\frac{dI'_{CEI}}{dt} = \frac{1}{L}(V_{AA} - I'_{CEI}R - V'_{A}) \tag{5.60}$$

where V_A is the device collector voltage $\cong V_{BC}$ (applied collector–base voltage of the bipolar transistor) for high-voltage operation. The current and voltage turn-off waveforms are derived [9] by integrating the above equation simultaneously with the state equations of IGBT, Eq. (5.58) and Eq. (5.59). The initial conditions are obtained from steady-state analysis. Using specific values for R, V_{AA}, and base lifetimes, the turn-off voltage waveforms are calculated for various inductance values. Comparison with the measured waveforms revealed that the model satisfactorily describes the transient waveforms for general switching conditions. Further details can be found in reference 9.

5.3.4 Simplified Model of *dV/dt* During Inductive Load Turn-off

The above model is a complex function of the redistribution and displacement capacitances, which are not readily amenable to calculations. A simple model has been proposed [11–13] for predicting the *dV/dt* of IGBT during the current boundary phase in which the device carries the full-rated current and the voltage across it rises to the bus voltage. Figure 5.8 shows the voltage

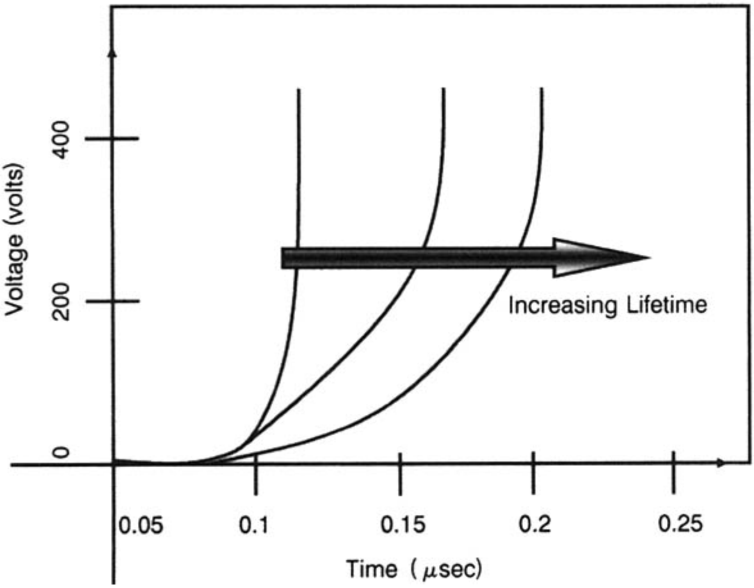


Figure 5.8 Typical turn-off waveforms of the IGBT during the current boundary phase.

waveforms for the current boundary phase during inductive load IGBT turn-off for different lifetimes. For the asymmetric IGBT structure, the carrier profile during the transient is approximated as a linear profile:

$$p(W_d) = p_0\left(1 - \frac{W_d}{d}\right) \tag{5.61}$$

where $p(W_d)$ is the charge in the drift region at the depletion layer edge, p_0 is the initial steady-state carrier distribution, W_d is the width of the depletion layer, and d is the drift region thickness. As the depletion boundary stretches to support the voltage, the excess carriers present in the drift region are swept out resulting in a current

$$I_T = qAp(W_d)\left(\frac{dW_d}{dt}\right) \tag{5.62}$$

where dW_d/dt is the rate of movement of the depletion boundary. It is expressed in terms of the voltage supported across the depletion layer as

$$\frac{dW_d}{dt} = \sqrt{\frac{\varepsilon_0 \varepsilon_s}{2qN_D V(t)}}\left(\frac{dV}{dt}\right) \tag{5.63}$$

where N_D is the dopant density of the drift region and $V(t)$ is the voltage supported across the depletion layer at time t. Combining Eqs. (5.62) and (5.63), we have

$$\frac{dV}{dt} = \frac{J_T}{qp(W_d)}\sqrt{\frac{2qN_D V(t)}{\varepsilon_0 \varepsilon_s}} \tag{5.64}$$

which reduces to the equation

$$\frac{dV}{dt} = \frac{J_T N_D d}{\varepsilon_0 \varepsilon_s p_0} \tag{5.65}$$

with the help of Eq. (5.61). This equation shows that the variation of voltage with time is linear during the current boundary phase. It is easy to see that dV/dt changes with lifetime because p_0 is lifetime-dependent, as shown in Fig. 5.8.

Example 5.8 In an IGBT, the dopant density of N^--drift region is 1×10^{14} cm^{-3}, and its thickness is 50 μm. The P-base junction depth is 3 μm and the depletion layer extends up to 3 μm into the drift region. The operational current density is 300 A/cm^2. Determine the voltage–time waveforms during inductive load turn-off for ambipolar lifetime values of 2, 5, and 10 μsec.

Table E.5.8.1 Variation of Turn-off Voltage with Time

Serial No.	Time (μsec)	Voltage for $\tau_a = 2$ μsec	Voltage for $\tau_a = 5$ μsec	Voltage for $\tau_a = 10$ μsec
1	0	0	0	0
2	0.05	22.06	20.21	19.58
3	0.1	44.11	40.41	39.15
4	0.15	66.17	60.62	58.73
5	0.2	88.23	80.83	78.3
6	0.25	110.28	101.03	97.88
7	0.3	132.34	121.24	117.45
8	0.35	154.4	141.44	137.03
9	0.5	220.57	202.1	195.8
10	0.75	330.85	303.1	293.63
11	1.0	441.13	404.13	391.51

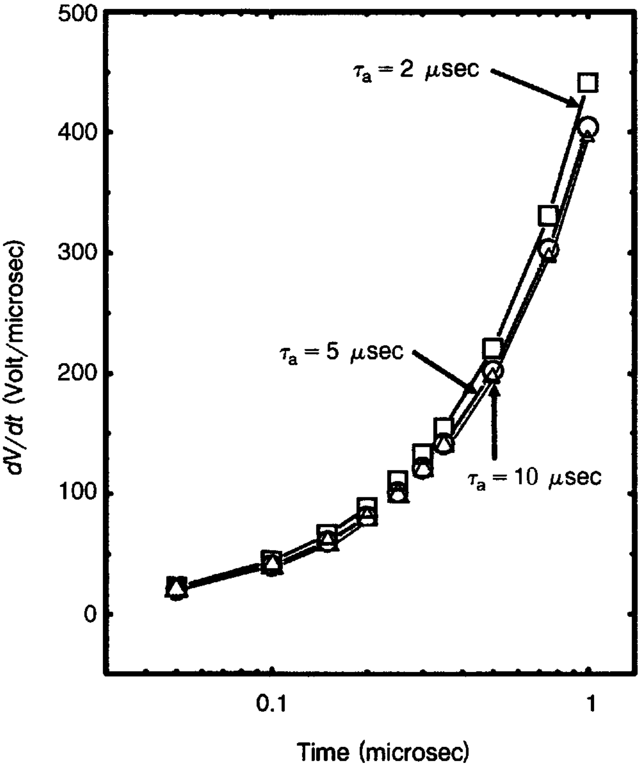


Figure E5.8.1 Calculated *dV/dt*-time waveforms for inductive load turn-off.

For $\tau_a = 2$ μsec, $L_a = \sqrt{(18.34 \times 2 \times 10^{-6})} = 6.0564 \times 10^{-3}$ cm. Hole concentration in the N$^-$-drift region at the junction between P$^+$ substrate and N$^-$-drift region is $p_0 = (300 \times 6.0564 \times 10^{-3})/(2 \times 1.6 \times 10^{-19} \times 12.43)[\sinh\{(50-3-3) \times 10^{-4}/(6.0564 \times 10^{-3})\}]/[\cosh\{(50-3-3) \times 10^{-4}/(6.0564 \times 10^{-3})\}] = 2.84 \times 10^{17}$ cm^{-3}. Proceeding in exactly the same manner for $\tau_a = 5$ μsec, we obtain $L_a = 9.576 \times 10^{-3}$ cm and $p_0 = 3.1 \times 10^{17}$ cm^{-3}. For $\tau_a = 10$ μsec, we obtain $L_a = 0.01354$ cm, and $p_0 = 3.2 \times 10^{17}$ cm^{-3}. Now, for $\tau_a = 2$ μsec, we get $dV/dt = (300 \times 1 \times 10^{14} \times 44 \times 10^{-4})/(8.854 \times 10^{-14} \times 11.9 \times 2.84 \times 10^{17} \times 1 \times 10^{6})$ V/μsec = 441.13 V/μsec. Likewise for $\tau_a = 5$ μsec, we obtain $dV/dt = 404.13$ V/μsec, and for $\tau_a = 10$ μsec, we get $dV/dt = 391.51$ V/μsec. The calculated values of voltages at different instants of time are listed in Table E5.8.1, and the same are plotted in Fig. E5.8.1.

One-dimensional non-quasi-static analytical models are presented [12] for describing the turn-off behavior of non-punchthrough IGBT (NPT-IGBT) driven in a chopper circuit with a clamped inductive load (Fig. 5.9), under the assumption of quasi-stationary conditions for the electron–hole plasma stored in the N$^-$ base. The turn-off process is analyzed by solving the ambipolar diffusion equation and the ambipolar current equations, neglecting the

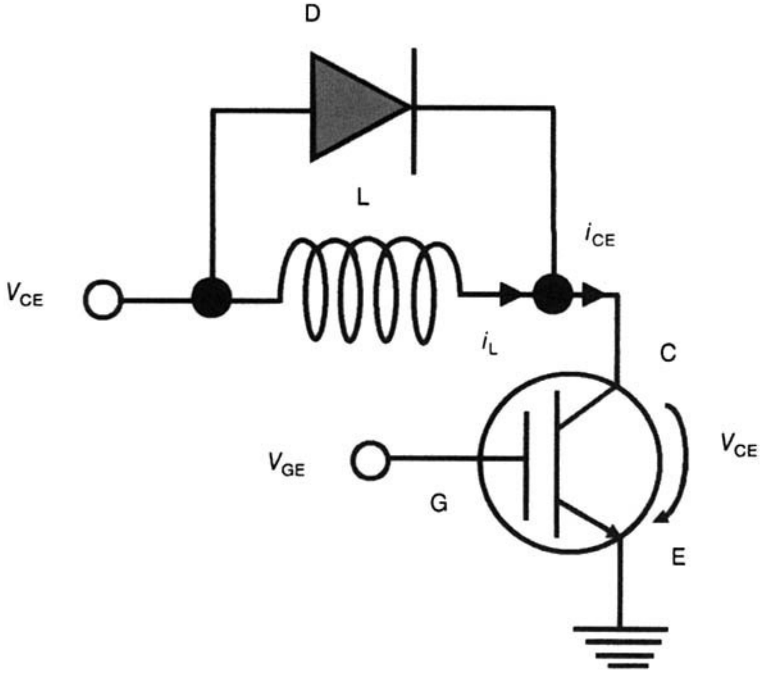


Figure 5.9 IGBT chopper circuit with clamped inductive load.

recombination effects because the time interval examined is only a few percent of the carrier lifetime in the N^- base. The ambipolar diffusion equation is solved by the integral method. In the quasi-neutral base regions where the carrier concentration alters rapidly with time, the carrier distribution is expressed as an nth-degree polynomial in space variables having time-dependent coefficients. These regions are referred to as the boundary layers, and the coefficients are evaluated by integrating the diffusion equation over the boundary layer, applying the charge conservation principle and physical considerations of the carrier profile in the remaining quasineutral part of N^- base. During the starting phase of turn-off, the accumulation layer near the gate disappears, due to the action of the field-driven current. This field-driven current is superimposed on the current components arising from the load current. The additional current extracts electrons from the semiconductor surface and transports holes toward it. During this extraction phase, the collector–emitter current remains constant, but the collector–emitter voltage rises. After this time domain, the tail phase commences in which the collector–emitter current begins to decay and the load current is increasingly forced through the diode. The electrons leaving the N^- base are removed from the boundary of the collector space-charge region, which expands accompanied by rise of collector–emitter voltage. The quasi-neutral N^--base region is divided into two parts, namely, the layer in which carrier extraction occurs and the one in which carrier concentration is maintained constant. During both the extraction phase and the early tail phase, analytical expressions are derived for the carrier profiles. The transient dependence of the collector–emitter voltage of the IGBT and the power losses in the extraction phase are obtained. It is demonstrated that the current decay mainly takes place due to the withdrawal of carriers from the boundary of the collector depletion region and that the collector–emitter current in the early tail phase is principally borne by electron flow.

Considering the *extraction phase* (Fig. 5.10a) during the action of the inductive load, the collector–emitter current remains constant at the value $i_{\mathrm{CE}} = i_{\mathrm{L}}$ because the displacement current through the diode is negligibly small. Furthermore, the ratio between the electron and hole currents at the P^+-collector/N^--base junction remains constant. Also, the absence of v_{GE} leads to withdrawal of electron support from the MOS segment of the IGBT. The net result is that electrons leaving the N^- base are extracted from the boundary of the collector–emitter depletion region. The accompanying enlargement of the depletion region causes rise of collector–emitter voltage v_{CE}. As shown in Fig. 5.10a, the quasi-neutral portion of the N^- base is subdivided into two parts, namely, a boundary layer $\varepsilon(t)$, which is subjected to carrier extraction, and the remaining part in which the carrier profile remains constant. The symbol $y_0(t)$ represents the position of the depletion region DR at time $t > 0$, and $y_{0\mathrm{T}}$ is its location at the end of the extraction phase $t = t_{\mathrm{T0}}$. The symbol p_ε denotes the carrier profile in the boundary

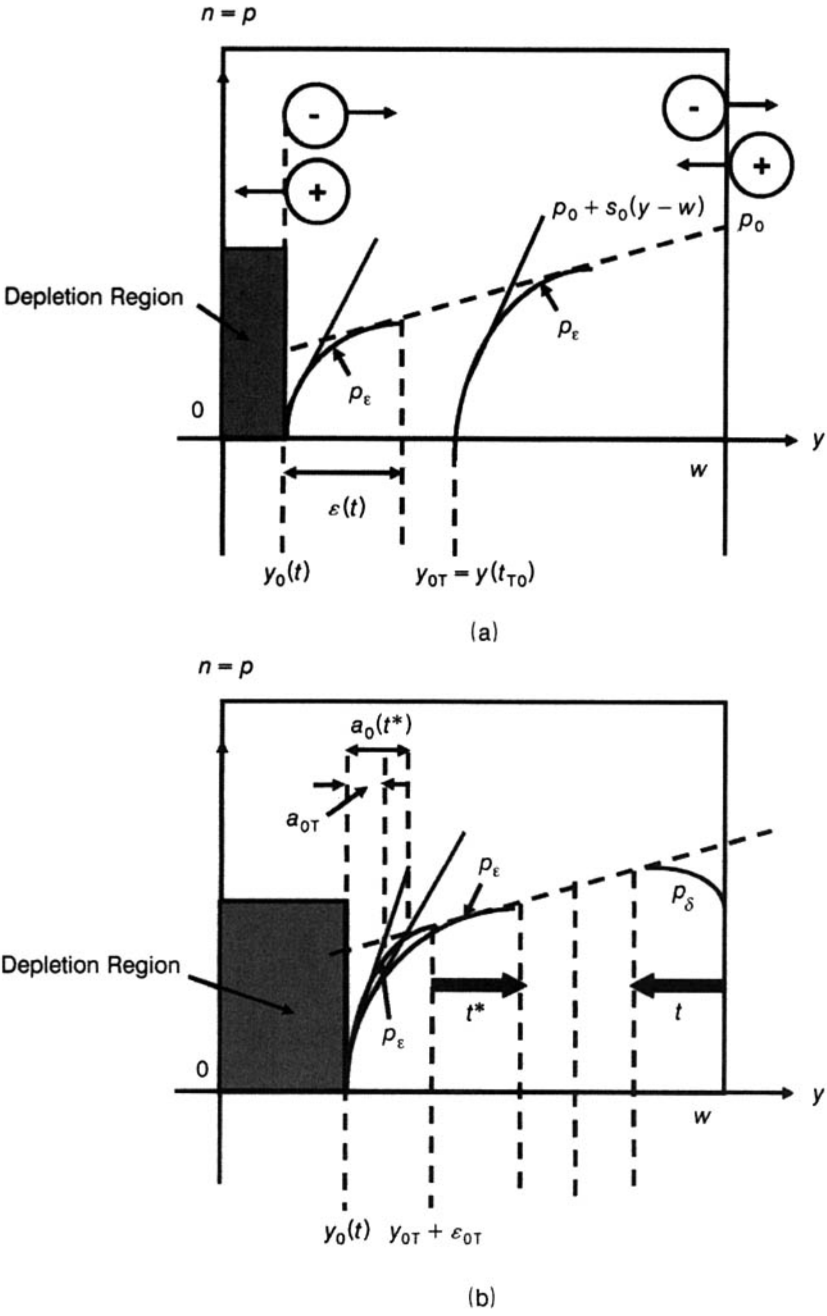

n = p
-
+
-
+
$p_0 + s_0(y - w)$
p_0
Depletion Region
p_ε
p_ε
0
y
w
$\varepsilon(t)$
$y_0(t)$
$y_{0T} = y(t_{T0})$
(a)
n = p
$a_0(t^*)$
a_{0T}
p_ε
p_δ
Depletion Region
t^*
t
p_ε
0
y
w
$y_0(t)$
$y_{0T} + \varepsilon_{0T}$
(b)

Figure 5.10 Carrier profiles and boundary layers in the N^- base of IGBT during the: (a) extraction phase and (b) early tail phase. **Notation:** $\varepsilon(t)$, Boundary layer of the quasi-neutral portion of N^- base from which carriers are removed; $\varepsilon(t^*)$ and $\delta(t^*)$, boundary layers; p_ε, carrier profile in the layer $\varepsilon(t)$; $p_0 + s_0(y - w)$, initial carrier profile at $t = 0$ for $y < y_0(t) + \varepsilon(t)$; $y_0(t)$, position of depletion region at $t = 0$; y_{0T}, position of depletion region at instant $t = t_{T0}$, $t = t_{T0}$, instant at which early tail phase begins; t^*, time scale ($t^* = t - t_{T0}$).

layer. The dashed line $\{p_0 + s_0(y - w)\}$ signifies the initial carrier profile at $t = 0$ for $y < y_0(t) + \varepsilon(t)$. The equation for s_0 is [12]

$$s_0 = \left.\frac{\partial p_\varepsilon}{\partial y}\right|_{y_0(t)+\varepsilon(t)} \tag{5.66}$$

This equation results from the requirements of continuity of the carrier profile at the location $y(t) + \varepsilon(t)$, and the existence of the derivative with respect to y. It is the actual carrier profile for $t > 0$ in the region $y \geq y_0(t) + \varepsilon(t)$ of the quasi-neutral base unaffected by the carrier extraction. Thus the parameters $y_0(t)$, $\varepsilon(t)$, and $p_\varepsilon(t)$ describe device behavior in the extraction phase. Assuming that the initial potential drop v_{CE} is zero, the dependence of voltage drop on time is expressed as

$$v_{\text{CE}}(t) = \frac{qN_{\text{D}}^* {y_0}^2(t)}{2\varepsilon_0 \varepsilon_{\text{s}}} \tag{5.67}$$

where qN_{D}^* is the charge density in the depletion region DR consisting of the background dopant density N_{D} and the hole concentration carrying the current density j_0, so that

$$N_{\text{D}}^* = N_{\text{D}} - \frac{j_0}{qv_{\text{psat}}} \tag{5.68}$$

where v_{psat} is the hole saturation velocity.

During the ensuing *early tail phase* (Fig. 5.10b), beginning at time $t = t_{\text{T0}}$ or at $t^* = 0$ in the time scale $t^* = t - t_{\text{T0}}$, there is a sharp decay of the collector–emitter current i_{CE} as the load current is increasingly forced through the diode, but the voltage v_{CE} is constant. The behavior of IGBT is determined by the simultaneous action of two phenomena: (i) Decay of collector–emitter current density $j_{\text{CE}}(t^*)$, primarily influenced by the carrier extraction at the boundary of the depletion region DR. (ii) Time-varying ratio of hole and electron currents at the collector, related to the extraction of carriers from the N$^-$-base near the N$^-$-base/P$^+$-collector junction. Thus two boundary layers $\varepsilon(t^*)$ and $\delta(t^*)$ approach toward each other with increasing time t^* until they eventually merge with each other, as illustrated in Fig. 5.10b.

The steep decay of current $j_{\text{CE}}(t^*)$ in the early tail phase is described by the simple equation

$$\frac{j_0(t^*)}{2qD_{\text{p}}s_0} = \frac{j_0/(2qD_{\text{p}}s_0)}{\varepsilon(t^*)/\varepsilon_{0\text{T}}} - 1 \tag{5.69}$$

where

$$\varepsilon(t^*)/\varepsilon_{0\mathrm{T}} = \sqrt{1 + 3.33\frac{b}{1+b}\frac{t}{a_{0\mathrm{T}}{}^2/2D_\mathrm{p}}} \tag{5.70}$$

Analysis of the transient development of electron, hole, and total current densities at the collector shows that in the early tail phase, electrons mainly carry the current. Furthermore, in the early tail phase, there is a possibility of extraction of holes out of the N^- base over the P^+ collector yielding a positive hole current density. From the direction of the electric field within the collector–emitter depletion region DR, the electron current density becomes zero at the boundary of DR. Thus the electrons can leave the N^- base at the metallurgical junction between N^- base and P^+ collector regions, resulting in a time-dependent electron current density. On the other hand, the holes are chiefly extracted from the N^- base by the total current density j flowing at the collector–emitter DR boundary as a pure hole current. If hole extraction differs from electron extraction, a hole current must flow across the above metallurgical junction for maintaining quasi-neutrality. The sign of this hole current is determined by the ratio $j_\mathrm{n}/j_\mathrm{p}$ at the junction. In the early tail phase, hole extraction at the metallurgical junction does occur, whereas for larger time intervals t^* the collector–emitter current is increasingly conducted by holes.

5.3.5 Dynamic Electrothermal Model of IGBT

This model [14–23] contains four terminals: three electrical terminals and one thermal terminal (Fig. 5.11). The electrical terminals are connected to the component models of the electrical network. Similarly, the thermal terminal is connected to the thermal network component models. The electrical model is a temperature-dependent model. It computes the instantaneous device parameters in terms of the IGBT chip surface temperature at that instant. It is based upon the temperature dependence of the IGBT model parameters, as well as the variation of Si properties with temperature. The thermal model determines the evolution of temperature distribution in the thermal network and thereby calculates the instantaneous Si surface temperature values, which are supplied to the electrical model. In this manner, the dynamics of the electrothermal interactions are accounted for. These simulations describe the IGBT chip surface temperature and the change in electrical characteristics on device heating.

The thermal network component models are obtained from the heat diffusion equation using the geometry of the component, material properties, and so on. To determine the thermal field in a medium (i.e., the temperature

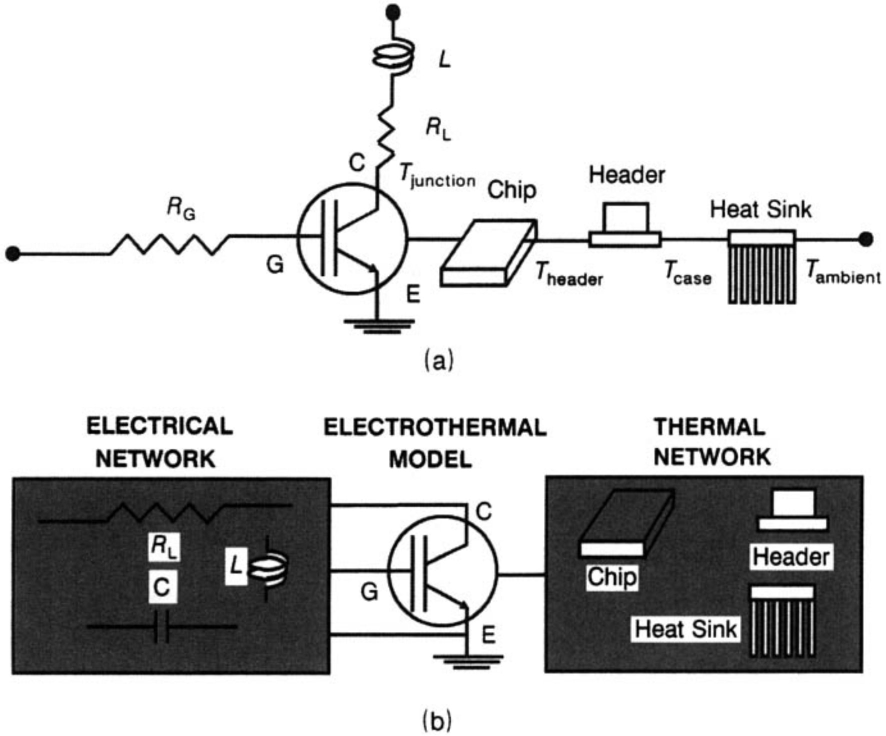

L
R_L
C
T_junction
Chip
Header
Heat Sink
R_G
G
E
T_header
T_case
T_ambient
(a)
ELECTRICAL NETWORK
ELECTROTHERMAL MODEL
THERMAL NETWORK
R_L
C
L
C
G
E
Chip
Header
Heat Sink
(b)

Figure 5.11 Interaction of electrical and thermal networks through electrothermal models for IGBT.

distribution), the heat diffusion equation in rectangular Cartesian coordinates is expressed as [24]

$$\frac{\partial}{\partial x}\left(k\frac{\partial T}{\partial x}\right)+\frac{\partial}{\partial y}\left(k\frac{\partial T}{\partial y}\right)+\frac{\partial}{\partial z}\left(k\frac{\partial T}{\partial z}\right)+\dot{q}=\rho c_{\mathrm{p}}\frac{\partial T}{\partial t} \tag{5.71}$$

where k is the thermal conductivity of Si given by $k(T)=1.5486(300/T)^{4/3}$, $\dot{q}$ is the rate of generation of energy per unit volume of the medium ($\mathrm{W/m^3}$), ρ is the mass density of the medium, and c_{p} is its specific heat. The product ρc_{p} ($\mathrm{J/m^3 \cdot K}$) is a measure of the ability of the material to store thermal energy. It is called the *volumetric heat capacity*.

The heat diffusion equation is a statement of the law of conservation of energy. Figure 5.12 illustrates a differential control volume $dx\,dy\,dz$ for conduction analysis. The conduction rates q_x, q_y, and q_z are the conduction heat rates perpendicular to the control surfaces at x, y, and z coordinate locations. The term $(\partial/\partial y)\{k(\partial T/\partial y\}$ = total conduction heat flux into the

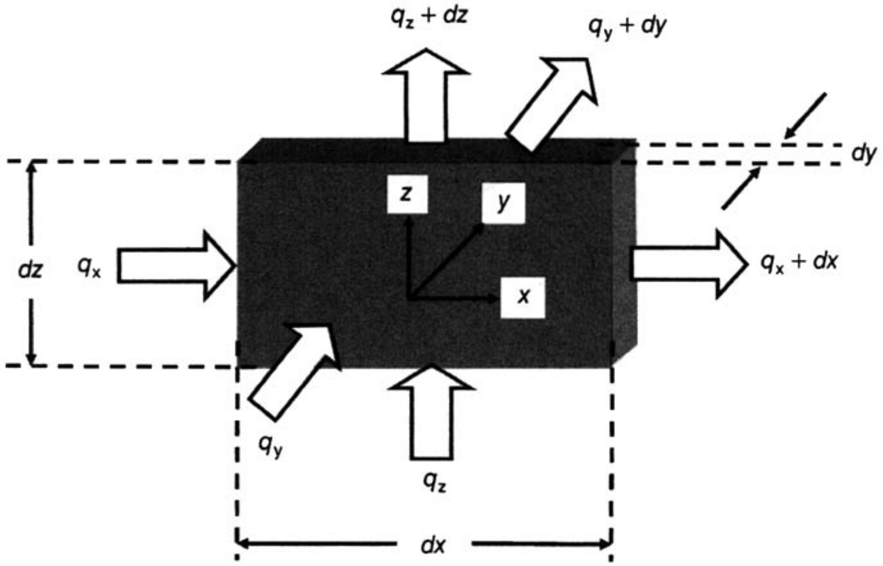


Figure 5.12 Differential control volume used for conduction analysis.

control volume for the *y*-coordinate. Hence,

$$\frac{\partial}{\partial y}\left(k\frac{\partial T}{\partial y}\right)dy = q''_y - q''_{y+dy} \tag{5.72}$$

where q'' denotes the heat flux. Applying this interpretation, the heat diffusion equation may be enunciated as follows: At any point in the medium, the rate of energy transfer by conduction into a unit volume together with the volumetric rate of generation of heat equals the rate of change of heat energy stored within the volume.

The major contribution to heat produced in a semiconductor device originates from the scalar product of the electric field E and the current density J. In the IGBT, the power dissipated consists of three components: (i) Power dissipated in the conductivity-modulated base resistance R_B and emitter–base depletion region, through carrier–lattice structure collisions, is

$$P_B = V_{EB0}I_{CEI} + R_B I_{CEI}^2 \tag{5.73}$$

where V_{EB0} is the emitter–base diffusion depletion potential. (ii) Power dissipated by the holes flowing through the collector–base depletion region is

$$P_{BC} = V_{BC}(I_{CEI} - I_{MOS}) \tag{5.74}$$

where I_{MOS} is the MOSFET current. (iii) Power dissipated by electrons flowing through the collector–base junction due to collisions with the lattice atoms:

$$P_{\mathrm{MOS}} = V_{\mathrm{BC}} I_{\mathrm{MOS}} \tag{5.75}$$

The sum-total power dissipation is obtained by adding together the three components:

$$P_{\mathrm{T}} = P_{\mathrm{B}} + P_{\mathrm{BC}} + P_{\mathrm{MOS}} \tag{5.76}$$

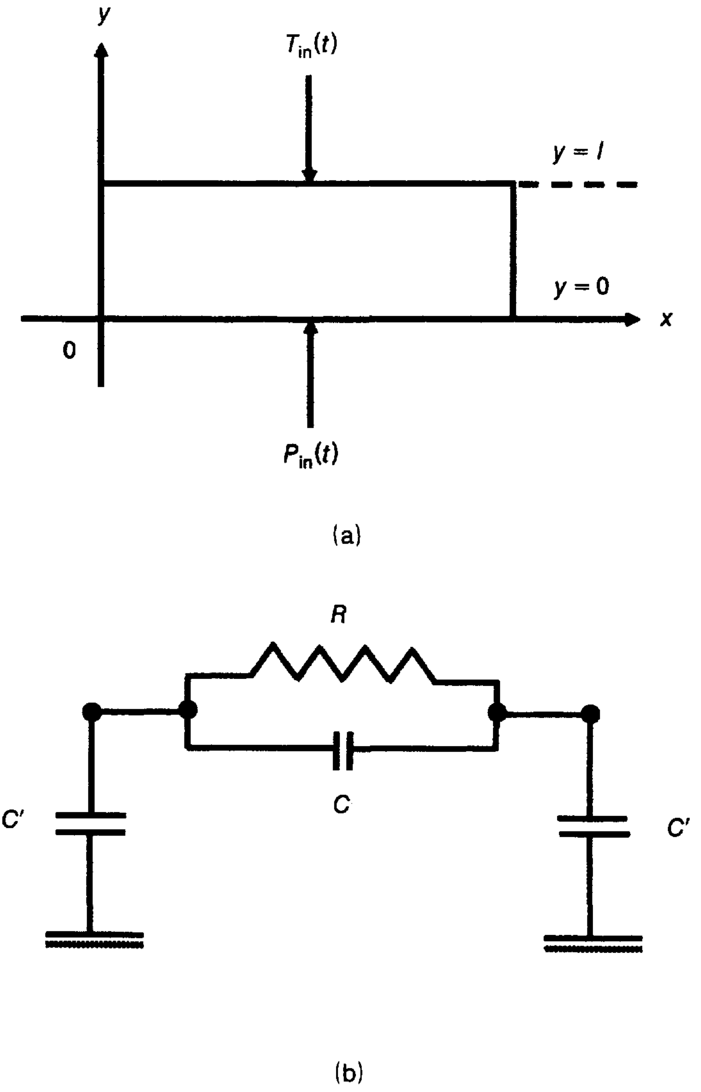


Figure 5.13 (a) One-dimensional chip representation for thermal modeling. (b) Unidimensional thermal network between two nodes.

The electrical model calculates this power dissipation. Then, applying the thermal model, the temperatures at various regions in the IGBT are determined. In commonly applied practice, the heat produced at the top Si surface is considered to flow uniformly, along the vertical y axis (Fig. 5.13a). In this framework, the top surface is a geometrical boundary of the device at $y = 0$, where the input power $P_{\text{in}}(t)$ is uniformly dissipated. The bottom surface at $y = l$ is the cooling boundary where the temperature is presumed to be T_{in}, the input temperature. Ignoring convection and radiation losses, unidimensional heat flow may be considered, giving

$$\frac{\partial}{\partial y}\left\{k(T)\frac{\partial T(t,y)}{\partial y}\right\} = \rho c \frac{\partial T(t,y)}{\partial t} \tag{5.77}$$

with the boundary conditions

$$Ak\left.\frac{\partial T}{\partial y}\right|_{y=0} = -P_{\text{in}}(t) \tag{5.78}$$

and

$$T(t, y = l) = T_{\text{in}}(t) \tag{5.79}$$

where A is the effective area of the semiconductor device. Because the semiconductor device models are implemented in circuit simulators, thermal circuit networks are the practical models for these simulations. The heat flow problem is simplified by using analogous electrical methods and representing the heat flow by the current flow in an electrical transmission line approximated by a lumped circuit consisting of resistances and capacitances (Table 5.1).

Thermal networks are constructed based on the finite element technique, equivalent to the discrete elements of the heat equation. A few advantages of the finite element method are as follows: (i) Conservation laws are exactly satisfied by coarse approximations. (ii) Irregular geometries are amenable to

Table 5.1 Relationship Between Electrical and Thermal Quantities

Serial No.	Electrical Quantity	Corresponding Thermal Quantity
1	Current source	Heat generator
2	Current (ampere)	Power dissipation (watt)
3	Electrical resistance (Ω)	Thermal resistance (°C/watt)
4	Electrical capacitance (coulomb/volt or farad)	Thermal capacitance (watt-sec/°C)
5	Potential difference (volts)	Temperature difference (°C)
6	Electrical impedance (volt/A or Ω)	Thermal impedance (°C/watt)

simple analysis. (iii) Implementation of local mesh refinement is straightforward. (iv) Construction of higher-order approximations is easy. A disadvantage of this method is the greater programming complexity.

The partial differential equation, for one-dimensional heat flow, is discretized into a finite number of first-order differential equations, assuming that temperature gradients and thermal conductivity do not vary significantly between neighboring points. The device under analysis is divided into small regions. The reference point is termed the *nodal point* or *node*, and the aggregate of points constitutes a *nodal network*, *grid*, or *mesh*. So, each node represents a particular region and its temperature is the average temperature of the region. The number and locations of thermal nodes within the component determine accuracy of the thermal model. If the number is large or mesh is fine, the numerical accuracy of the calculations is high. For a steady-state problem, the heat balance requires that quantity of heat entering each node must equal the quantity of heat leaving the node.

The semidiscrete equation derived from a *Galerkin* finite-element projection is given by [20]

$$\mathbf{M}\dot{d} + K\mathbf{d} = \mathbf{F} \tag{5.80}$$

where **M** is the mass matrix, **K** is the stiffness matrix, **F** is the force vector, $\mathbf{d}(t)$ is the vector containing all nodal temperatures d_i's, and $\mathbf{d}(t)$ is the derivative of $d(t)$ with respect to time. **M** is realized by capacitive elements, **K** by resistors, and **F** by current sources. In this manner, thermal effects are incorporated into circuit simulators. The one-dimensional thermal network is shown in Fig. 5.13b. Here the parameters are as follows (Appendix 5.7):

$$R = \frac{\delta_x}{kA}, \qquad C = \frac{A\,\delta_x\,\rho c_p}{2}, \qquad C' = \frac{A\,\delta_x\,\rho c_p}{6} \tag{5.81}$$

where δ_x is the discretization step, k is the thermal conductivity of Si, ρ is the density, c_p is the specific heat, and A is the cross-sectional area.

Thermal capacitance is the product of the mass of the material and its specific heat. If on absorbing a quantity of heat Δq, the temperature of a specimen rises by ΔT, its thermal capacitance $C = \Delta q/\Delta T$. Any increase of thermal capacitance will cause the concerned region to respond more slowly to environmental temperature changes, increasing the time required to attain equilibrium. Thermal capacitances allow for the time-dependent or transient effects. For transient problems, the amount of heat entering the node minus that leaving the node during a small interval of time should be equal to Thermal capacitance of the node × Rise of nodal temperature during the time interval considered.

Now by application of the electrical model, the different physical and electrical parameters at the evaluated temperatures are found. Several competing mechanisms come into play to produce the observed temperature

Table 5.2 Temperature Dependency of Selected Parameters Used in Electrothermal Simulations

Serial No.	Physical Parameter	Symbol	Equation	Unit
1	Thermodynamic voltage	ϕ_0	$\phi_0 = \frac{kT}{q}$	V
2	Intrinsic carrier concentration	$n_i(T)$	$n_i(T) = 3.87 \times 10^{16} T^{3/2} \times \exp\left(-\frac{7.02 \times 10^3}{T}\right)$	cm^{-3}
3	Electron mobility	$\mu_n\,(T, N_D)$	$\mu_n(T,\ N_D) = 88(T/300)^{-0.57} + 1 + \frac{7.4 \times 10^8 T^{-2.33}}{1 + \left\{\frac{N_D}{1.26 \times 10^{17}(T/300)^{2.4}}\right\}^{0.88} (T/300)^{-0.146}}$	$\mathrm{cm}^2/\mathrm{V\text{-}sec}$
4	Hole mobility	$\mu_p\,(T, N_A)$	$\mu_p(T, N_A) = 54.3(T/300)^{-0.57} + 1 + \frac{1.36 \times 10^8 T^{-2.33}}{1 + \left\{\frac{N_A}{2.35 \times 10^{17}(T/300)^{2.4}}\right\}^{0.88} (T/300)^{-0.146}}$	$\mathrm{cm}^2/\mathrm{V\text{-}sec}$
5	Electron saturation velocity	$\upsilon_{n,\mathrm{sat}}(T)$	$\upsilon_{n,\mathrm{sat}}(T) = 1.434 \times 10^9 T^{-0.87}$	cm/sec
6	Hole saturation velocity	$\upsilon_{p,\mathrm{sat}}(T)$	$\upsilon_{p,\mathrm{sat}}(T) = 1.624 \times 10^8 T^{-0.52}$	cm/sec
7	Electron and hole lifetime	$\tau_{n,p}(T)$	$\tau_{n,p}(T) = \tau_{n,p}(300)(T/300)^{1.7}$	sec

8	Threshold voltage	$V_{\mathrm{Th}}(T)$	$V_{\mathrm{Th}}(T) = V_{\mathrm{FB}} + 2\psi_{\mathrm{B}}(T) + \sqrt{\dfrac{4\varepsilon_0\varepsilon_s N_{\mathrm{A}}\psi_{\mathrm{B}}(T)}{C_{\mathrm{ox}}}}$ where the most temperature-sensitive parameter is $\psi_{\mathrm{B}}(T) = \dfrac{kT}{q}\ln\left\{\dfrac{N_{\mathrm{A}}}{n_{\mathrm{i}}(T)}\right\}$	V
9	MOSFET current	$I_{\mathrm{MOS}}(T)$	$I_{\mathrm{MOS}}(T) = \begin{cases} 0, V_{\mathrm{GS}} < V_{\mathrm{Th}} \\ K_{\mathrm{plin}}\dfrac{(V_{\mathrm{GS}} - V_{\mathrm{Th}})V_{\mathrm{DS}} - \dfrac{K_{\mathrm{plin}}}{2K_{\mathrm{psat}}}V_{\mathrm{DS}}{}^2}{1 + \theta(V_{\mathrm{GS}} - V_{\mathrm{Th}})}, \\ V_{\mathrm{DS}} \le (V_{\mathrm{GS}} - V_{\mathrm{Th}})\dfrac{K_{\mathrm{psat}}}{K_{\mathrm{plin}}} \\ K_{\mathrm{psat}}\dfrac{(V_{\mathrm{GS}} - V_{\mathrm{Th}})^2}{2\{1 + \theta(V_{\mathrm{GS}} - V_{\mathrm{Th}})\}}, \\ V_{\mathrm{DS}} \ge (V_{\mathrm{GS}} - V_{\mathrm{Th}})\dfrac{K_{\mathrm{psat}}}{K_{\mathrm{plin}}} \end{cases}$ where θ is the transverse field transconductance factor (measured in V^{-1}), K_{plin} is linear region MOSFET transconductance parameter ($\mathrm{A/V}^2$), K_{psat} is saturation region MOSFET transconductance parameter, and $\dfrac{K_{\mathrm{plin}}}{K_{\mathrm{psat}}} = 1 + \sqrt{\dfrac{q\varepsilon_0\varepsilon_s N_{\mathrm{A}}}{\psi_{\mathrm{B}}}} \Big/ C_{\mathrm{ox}},$ $K_{\mathrm{plin}} = K_{\mathrm{plin\,300}}\left\{\dfrac{\mu_{\mathrm{n}}(T)}{\mu_{\mathrm{n}}(300)}\right\}$	A

dependence of electrical characteristics of an IGBT. For instance, considering the fall of saturation current with increasing temperature, we note that the underlying causes are the decrease in threshold voltage and the MOSFET transconductance parameter with rise of temperature. The temperature of the MOSFET channel (T_{Ch}) gives V_{Th} and μ_{ns}. The marginal reduction of bipolar transistor current gain with increasing temperature gives an additional contribution. Base transport factor is not significantly influenced by temperature because of oppositely directed changes of carrier lifetimes and diffusion coefficients. However, there is a slight reduction of emitter injection efficiency with increasing temperature because the decrease in the ratio $I_{\mathrm{s}}/n_{\mathrm{i}}$ is not as large as in diffusivity. Other examples using temperature distribution in the device are as follows: Temperatures in the collector–base depletion layer yield current values for thermal generation. Base region temperature leads to conductivity-modulated base resistance R_{B}. Temperature in the emitter–base depletion layer is used for calculating the emitter-base diffusion depletion potential. Useful formulae for temperature dependency of parameters in thermal model are compiled in Table 5.2.

Example 5.9 Write and solve the rate equation for heat conduction in a plane wall with air on both sides. Draw the thermal equivalent circuit and determine the total resistance.

A 1-cm × 1-cm IGBT chip is soldered on a 2-mm-thick copper substrate of equal area. The solder joint thickness is 0.1 mm. Both the chip and the substrate surfaces are air-cooled. The ambient temperature is 25°C and the convection coefficient of air is 100 $\mathrm{W/m^2}$. Also the thermal conductivity of Cu is 401 W/m-K and representative maximum thermal resistance of solid (Si)/solid (Cu) interface is 0.9×10^{-4} $\mathrm{m^2\ K/W}$. Derive the energy balance equation for the chip. Will the chip operate below a maximum permissible temperature of 125°C if it dissipates 10^4 $\mathrm{W/m^2}$ under normal conditions?

The rate equation for heat conduction (energy diffusion) is Fourier's law, which for a one-dimensional plane wall, Fig. E5.9.1a, having a temperature distribution $T(y)$ is expressed as $q_y'' = -k(dT/dy)$ where q_y'' is the heat flux (unit $\mathrm{W/m^2}$) = heat transfer rate per unit area in the y direction perpendicular to the direction of transfer. The proportionality constant k is the thermal conductivity of the wall material. The negative sign indicates that the heat transference occurs in the direction of decreasing temperature.

The basic tool for thermal conduction analysis is the heat diffusion equation. Under steady-state, one-dimensional conditions with no energy generation, the thermal flux is constant in the direction of heat flow, so that $(d/dy)(k\,dT/dy) = 0$. Integrating this equation twice, the general solution is $T(y) = C_1 y + C_2$ with the constants C_1 and C_2 determined from the boundary conditions: at $y = 0$, $T(0) = T_1$ and at $y = L$, $T(L) = T_2$. Then $T(y) = (T_2 - T_1)y/L + T_1$. This gives $dT/dy = kA(T_2 - T_1)/L$ and applying Fourier's law, $q_y = kA(T_1 - T_2)/L$. This equation shows that the heat rate q_y is a constant, independent of y. Therefore q_y is constant throughout the network.

For conceptualizing and quantifying heat transfer problems, the equivalent thermal circuit for plane wall is given in Fig. E5.9.1b. From the analogy of association of electrical resistance with the conduction of electricity, a thermal resistance is associated with the conduction of heat. Similarly, a thermal resistance is also used to

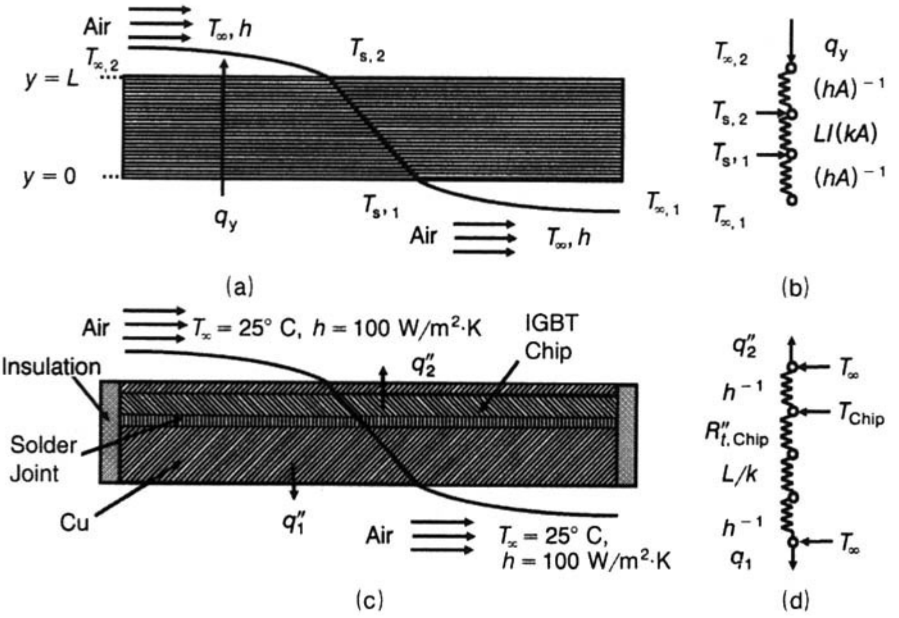

Air
T_∞, h
$T_{\infty,2}$
$y = L$
$y = 0$
$T_{s,2}$
q_y
$T_{s,1}$
Air
T_∞, h
$T_{\infty,1}$
$T_{\infty,2}$
$T_{s,2}$
$T_{s,1}$
$T_{\infty,1}$
q_y
$(hA)^{-1}$
$L/(kA)$
$(hA)^{-1}$
(a)
(b)
Air
$T_\infty = 25°$ C, $h = 100$ W/m^2·K
IGBT Chip
Insulation
q''_2
Solder Joint
Cu
q''_1
Air
$T_\infty = 25°$ C, $h = 100$ W/m^2·K
q''_2
h^{-1}
$R''_{t,\text{Chip}}$
L/k
h^{-1}
q_1
T_∞
T_{Chip}
T_∞
(c)
(d)

Figure E5.9.1 (a) Conduction through a plane wall; (b) its equivalent thermal circuit diagram; (c) conduction analysis through IGBT chip; and (d) equivalent thermal circuit for this system.

represent the heat transference by convection. As the conduction and convection phenomena take place sequentially, the conduction and convection thermal resistances are connected in series and may be simply added to obtain the total resistance.

The heat transfer rate is determined separately by each component in the network. Because q_y is constant throughout the network as shown above, we can write $q_y = (T_\infty - T_1)/\{1/(hA)\} = (T_1 - T_2)/\{l/(kA)\} = (T_2 - T_\infty)/\{1/(hA)\}$. Due to the series connection of conduction and convection resistances, as already mentioned, the total resistance is given by $R_{\text{tot}} = 1/(hA) + L/(kA) + 1/(hA) = 2/(hA) + L/(kA)$.

Figure E5.9.1c shows the schematic of the IGBT chip and Fig. E5.9.1d gives its equivalent thermal circuit. It is assumed that the chip has negligible thermal resistance so that an isothermal condition exists in the chip. Furthermore, the chip assembly is insulated from the sides so that heat transference takes place only along the y direction. Also, radiation exchange with the surroundings is very small. Thus, heat dissipated in the IGBT chip is transferred directly to the air from the exposed surface and indirectly through the solder joint and Cu substrate. Performing an energy balance analysis, we have $q''_{\text{chip}} = q''_1 + q''_2$ or $q''_{\text{chip}} = (T_{\text{chip}} - T_\infty)/(1/h)(T_{\text{chip}} - T_\infty)/\{R''_{\text{t,Chip}} + (L/k) + (1/h)\}]^{-1}$. For a worst-case estimation of T_{chip}, the maximum possible value of 0.9×10^{-4} K/W is taken. Then $T_{\text{chip}} = T_\infty + q''_{\text{chip}}[h + 1/\{R''_{\text{t, Chip}} + (L/k) + (1/h)\}]^{-1} = 25°\text{C} + 10^4\ \text{W/m}^2 \times [100 + 1/\{(0.9 + 0.05 + 100) \times 10^{-4}]^{-1}\ \text{m}^2.\text{K/W} = 25°\text{C} + 50.24°\text{C} = 75.24°\text{C}$. Thus the chip will operate below its maximum allowable temperature.

It would be worthwhile to point out here that like any power semiconductor device, there are two destruction modes of IGBTs, namely, thermal and electrical [25, 26]. *Thermal Destruction* is activated by the disappearance of the built-in potential between the N^+ emitter and P base of the integrated DMOSFET in the IGBT. This potential disappears thermally, with a coefficient of 2 mV/K. Starting from a value of 0.65 V at 300 K, this potential vanishes at 650 K because at this temperature the built-in potential decreases by an amount $= 350 \times 2 = 650$ mV. The result is thermal runaway and destruction of the device. Thus thermal destruction takes place at a temperature of 650 K, as soon as the above-mentioned potential decreases to zero. Numerical analysis and experimental investigations have corroborated this fact.

Electrical destruction occurs when the power dissipation of the IGBT chip exceeds a critical value ~ 2000 kW/cm^2. This type of destruction is initiated by impact ionization and avalanche injection across the N^--drift region/N^+-buffer layer and N^--drift region/P-base junctions of the IGBT. The resulting regenerative feedback mechanism of increasing current of the PNP transistor is responsible for device destruction, through loss of gate control. The critical power dissipation of 2000 kW/cm^2 for the IGBT is an order of magnitude higher than that for the bipolar transistor. It also depends on the h_{FE} of the PNP transistor.

5.3.6 Extraction of Model Parameters for Circuit Analysis

The physics-based IGBT model of static and dynamic behavior has been implemented in the IG-Spice or Saber circuit simulator software [27–29]. The first step in applying a physics-based IGBT model for circuit analysis consists in characterization of the actual IGBT to be used in the circuit, through a series of static and dynamic measurements, namely, output characteristics (I_{CE}–V_{CE} for different V_{GE}), transfer characteristics (I_{CE}–V_{GE} for different V_{CE}), gate charging waveforms (V_{CE} and V_{GE} with respect to time), transient studies, and capacitance–voltage (C–V) records. Measurement results are fed into a software package, which employs an optimization algorithm to perform the best fit between the experimental data and IGBT parameters. Thus the parameters are reconstructed from the measurements without assuming any knowledge of the internal IGBT details [30].

5.4 CONCLUDING COMMENTS

Two-dimensional effects significantly influence the ON-state characteristics of the device, and due attention must be paid to them in IGBT modeling. It is reemphasized that physical understanding of IGBT operation and modeling of electrical behavior of the device will pave the way toward successful IGBT design. A thorough grasp of the subject matter of this chapter will ensure a smooth ride ahead. In the forthcoming chapter, we address the vital issue of immunizing IGBTs against latchup to provide reliable device operation.

REVIEW EXERCISES

5.1 In the cross-sectional diagram of IGBT, identify the MOSFET and PIN rectifier sections. Hence justify that the IGBT can be treated as a PIN rectifier in series with a MOSFET.

5.2 Why is the current density in the PIN rectifier of IGBT approximately equal to the collector current density?

5.3 Interrelate the current I_{MOSFET} flowing in the channel of IGBT, the collector current density J_C, unit cell width W of the cross section and the cell width Z perpendicular to the cross section.

5.4 In an IGBT operating in its forward conduction mode, sufficient gate bias is applied in order that the potential drop across the IGBT is low. Explain whether under this condition the MOSFET section of IGBT is working in its linear region or saturation region?

5.5 Does the forward conduction current density of IGBT rise linearly or exponentially with forward bias voltage? Explain giving the necessary equation.

5.6 Using the ON-state current density and potential drop as the deciding factors, explain why a smaller chip size is used for IGBT as compared to a power MOSFET and bipolar transistor?

5.7 Why is the transconductance of an IGBT higher than that of the MOSFET having the same cell size and channel length? For a typical common-base current gain value $(\alpha_{\text{PNP}}) = 0.5$, what is the ratio of transconductance of IGBT and MOSFET of equal channel aspect ratio?

5.8 Why does the carrier distribution in the N^--drift region of IGBT differ from that in a PIN rectifier? How does this difference change the boundary conditions for analyzing the IGBT structure? Mention the appropriate boundary conditions for IGBT.

5.9 Illustrate on a diagram the difference of free carrier distribution between the PIN rectifier and IGBT. How does this difference affect the conductivity modulation in IGBT, producing a large potential drop in its JFET region?

5.10 Discuss how the PIN rectifier–DMOSFET model of IGBT is useful for understanding the IGBT operation both in the linear and saturation regions? How does this model help in describing the forward conduction characteristics of IGBT as a function of carrier lifetime in the drift region? What is the impact of increasing the breakdown voltage on the forward characteristics of IGBT?

5.11 Does the PIN rectifier–DMOSFET model of IGBT take into account the hole component of current flowing into the P-base region of the device? How is this component taken into consideration in the bipolar transistor–DMOSFET model of IGBT? Write the relationship between the electron and hole components (I_n and I_p) and the current gain α_{PNP} of the PNP transistor in IGBT.

5.12 Give the equation for the forward conduction characteristics of IGBT showing how it accounts for the amplification of output current by PNP transistor.

5.13 In the analysis of IGBT under high-level injection conditions, why must ambipolar parameters for lifetime and diffusion length be used?

5.14 Give two reasons why the non-quasi-static analysis is necessary for describing the transient current and voltage waveforms of the IGBT. Are the non-quasi-static effects important for the high-gain, low-injection level operation of signal transistors?

5.15 Write the time-dependent ambipolar diffusion equation for the IGBT. What are the boundary conditions for the excess carrier concentration at the emitter and collector edges of the quasi-neutral base ($x = 0, W$)?

5.16 Give the equation expressing the time rate of change of quasi-neutral base width (dW/dt) in terms of the rate of change of collector base voltage (dV_{BC}/dt). Does the moving quasi-neutral base width boundary condition produce non-quasi-static current?

5.17 Describe how the analytical IGBT model is used in conjunction with the load circuit equations for modelling the device–circuit interactions of IGBT.

5.18 Why must two-dimensional effects be taken into account during forward conduction analysis of an IGBT? How does the 2-D analysis help in achieving superior ON-state characteristics?

5.19 Develop a simple analytical model for dV/dt of an IGBT during inductive load turn-off. How does this model show variation of dV/dt with carrier lifetime in the drift region?

5.20 Write the semidiscrete heat equation derived from finite element projection. Explain the different symbols used.

5.21 Enlist the major sources of power dissipation in IGBT. Give the mathematical equations describing them.

5.22 How is the IGBT chip represented in one dimension for thermal modeling? Draw and explain the 1-D thermal network. What is meant by thermal capacitance?

5.23 Cite some examples of electrical parameters of IGBT, which are sensitive to temperature changes. Write the equations for thermal behaviour of two such parameters.

5.24 Explain how the electrothermal interactions are taken into consideration dynamically in IGBT modelling? Has the physics-based IGBT model been implemented in a circuit simulator software?

5.25 The resistivity of the N^--drift region of an IGBT is 28 Ω-cm in an IGBT. Its thickness is 40 μm and P-base junction depth X_{jP} is 2.5 μm. If the ambipolar diffusion length in this region is 27 μm, determine the sum-total voltage drop across the P^+N junction and the N^--drift region of the IGBT working at a forward voltage of 1.9 V and current density of 90 A/cm^2.

5.26 The structural parameters of a linear cell geometry IGBT are as follows: width of N^--drift region $(d) = 45$ μm, channel length (L_{Ch}) is 1 μm, gate oxide thickness (t_{ox}) is 500 Å, distance between adjoining P-base wells (W_R) is 5 μm, total cell width $(W_R + W_P)$ is 27 μm. Its threshold voltage is 3.8 V, operating gate bias is 15 V, and current gain of PNP transistor $\alpha_{PNP} = 0.48$. Also, it is given that ambipolar diffusion coefficient D_a is 18 cm^2/sec, the function $F(d/L_a)$ is 0.2, and electron mobility in the surface inversion layer μ_{ns} is 400 cm^2/V-sec. Apply the first-order PIN rectifier–DMOSFET and bipolar transistor–DMOSFET models of IGBT to determine the forward voltage drop at current densities of 0.5, 5, 50, and 500 A/cm^2, and plot the collector–emitter current density–voltage characteristics of the IGBT.

5.27 Channel length of an IGBT is 0.8 μm, and its channel width is 152 cm. The drift region thickness is 158 μm. The carrier lifetime is 0.7 μsec, threshold voltage is 4.3 V, and oxide capacitance per unit area is 1.25×10^{-8} F/cm^2. The IGBT is working in the saturation region. Find the collector current and transconductance of the device at a gate voltage of 10.5 V. The electron mobility in the inversion layer is 500 cm^2/V-sec, and ambipolar diffusion coefficient is 18 cm^2/sec.

REFERENCES

1. B. J. Baliga, *Power Semiconductor Devices*, PWS Publishing Co., Boston, 1996.
2. W. Feiler, W. Gerlach, and U. Wiese, Two-Dimensional Models of the Carrier Distribution in the ON-State of the IGBT, *Solid State Electron.*, Vol. 38, No. 10, October 1995, pp. 1781–1790.
3. K. J. Binns and P. J. Lawrenson, *Analysis and Computation of Electric and Magnetic Field Problems*, 2nd edition, Pergamon Press, Oxford, 1973, p. 117.
4. E. Weber, *Electromagnetic Fields—Theory and Applications*, Vol. I, *Mapping of Fields*, Wiley, New York, 1960.
5. K. Sheng, S. J. Finney, and B. W. Williams, A New Analytical IGBT Model with Improved Electrical Characteristics, *IEEE Trans. Power Electron.*, New York, Vol. 14, No. 1, January 1999, pp. 98–107.
6. A. R. Hefner, Jr., and D. L. Blackburn, A Performance Trade-off for the Insulated Gate Bipolar Transistor: Buffer Layer versus Base Lifetime Reduction, *IEEE Trans. Power Electron.*, Vol. PE-2, No. 3, July 1987, pp. 194–207.
7. A. R. Hefner, Jr., and D. Blackburn, An Analytical Model for the Steady State and Transient Characteristics of the Power Insulated Gate Bipolar Transistor, *Solid State Electron.*, Vol. 31, No. 10, October 1988, pp. 1513–1532.
8. A. R. Hefner, Jr., Analytical Modelling of Device-Circuit Interactions for the Power Insulated Gate Bipolar Transistor (IGBT), *IEEE Trans. Industrial Applications*, Vol. 26, No. 6, November/December 1990, pp. 995–1005.
9. A. R. Hefner, Jr., An Improved Understanding for the Transient Operation of the Power Insulated Gate Bipolar Transistor (IGBT), *IEEE Trans. Power Electronics*, Vol. 5, No. 4, October 1990, pp. 459–468.
10. A. R. Hefner, An Investigation of the Drive Circuit Requirements for the Power Insulated Gate Bipolar Transistor, *IEEE Trans. Power Electron.*, Vol. 6, March 1991, pp. 208–219.
11. A. Ramamurthy, S. Sawant, and B. J. Baliga, Modelling the *dV/dt* of the IGBT During Inductive Turn-off, *IEEE Trans. Power Electron.*, Vol. 14, No. 4, July 1999, pp. 601–606.

12. W. Feiler, W. Gerlach, and U. Wiese, On the Turn-off Behaviour of the NPT-IGBT under Clamped Inductive Loads, *Solid State Electronics*, Vol. 39, No. 1, January 1996, pp. 59–67.
13. M. Trivedi and K. Shenai, Modelling the Turn-off of IGBTs in Hard- and Soft-Switching Power Converters, *IEEE Trans. Electron Devices*, Vol. 44, No. 5, May 1997, pp. 887–893.
14. A. R. Hefner, Jr., A Dynamic Electro-Thermal Model for the IGBT, *Conference Record of the IEEE Industry Applications Society (IEEE-IAS) Annual Meeting*, October 4–9, 1992, Houston, Texas, IEEE, New York, pp. 1094–1104.
15. V. Axelrad and R. Klein, Electrothermal Simulation of an IGBT, in *Proceedings 1992 International Symposium on Power Semiconductor Devices and ICs (ISPSD'92)*, Tokyo, Japan, IEEE, New York, 1992, pp. 158–169.
16. A. R. Hefner and D. L. Blackburn, Simulating the Dynamic Electrothermal Behaviour of Power Electronic Circuits and Systems, *IEEE Trans. Power Electron.*, Vol. 8, No. 4, October 1993, pp. 376–385.
17. P. Turkes, W. Kiffe, and R. Kuhnert, Critical Switching Conditions of a Non-punchthrough IGBT Investigated by Electrothermal Circuit Simulation, in *Proceedings of the 6th International Symposium on Power Semiconductor Devices and ICs, ISPSD'94, Davos, Switzerland*, May 31, June 2, 1994, IEEE, New York, pp. 51–55.
18. B. Fatemizadeh, D. Silber, M. Fullmann, and J. Serefin, Modelling of LDMOS and LIGBT Structures at High Temperatures, *Proceedings of the 6th International Symposium on Power Semiconductor Devices and ICs, ISPSD'94, Davos, Switzerland*, May 31, June 2, 1994, IEEE, New York, pp. 137–142.
19. A. Amimi, R. Bouchakour, and T. Maurel, Modelling and Simulation of Self-Heating and/or Degradation Effects for the IGBT Transistor, in *Proceedings of the PESA'96, Symposium on Modelling, Analysis, and Simulation*, Lille, France, July 9–12, 1996, pp. 855–860.
20. J. T. Hsu and L. Vu-Quoc, A Rational Formulation of Thermal Circuit Models for Electrothermal Simulation—Part I: Finite Element Method, *IEEE Trans. Circuits & Systems—I: Fundamental Theory and Applications*, Vol. 43, No. 9, September 1996, pp.721–732.
21. A. Ammous, B. Allard and H. Morel, Transient Temperature Measurements and Modelling of IGBTs Under Short-Circuit, *IEEE Trans. Power Electron.*, Vol. 13, No. 1, January 1998, pp. 12–25.
22. A. Ammous, K. Ammous, H. Morel, B. Allard, D. Bergogne, F. Sellami, and J. P. Chante, Electrothermal Modelling of IGBTs: Application to Short-Circuit Conditions, *IEEE Trans. Power Electron.*, Vol. 15, No. 4, July 2000, pp. 778–790.
23. R. Kraus, K. Hoffmann, and P. Turkes, Analysis and Modelling of the Technology-Dependent Electrothermal IGBT Characteristics, *Proceedings of the IPEC'95*, Yokohama, Japan, pp. 1128–1133.
24. F. P. Incropera and D. P. DeWitt, *Fundamentals of Heat and Mass Transfer*, fourth edition, John Wiley & Sons, New York, 1996, pp. 52–55, 74–77.
25. N. Iwamuro, A. Okamoto, S. Tagami, and H. Motoyama, Numerical Analysis of Short-Circuit Safe Operating Area for P-Channel and N-Channel IGBTs, *IEEE Trans. Electron Devices*, Vol. 38, No. 2, February 1991, pp. 303–309.
26. H. Hagino, J. Yamashita, A. Uenishi, and H. Haruguchi, An Experimental and Numerical Study of the Forward-Biased SOA of IGBTs, *IEEE Trans. Electron Devices*, Vol. 43, No. 3, March 1996, pp. 490–500.
27. F. F. Protiwa, O. Apeldoorn, and N. Gross, New IGBT Model for PSPICE, in *Proceedings of the 5th European Conference on Power Electronics and Applications EPE'93*, 13–16 September 1993, Brighton, U.K., IEE, London, pp. 226–231.
28. C. S. Mitter, A. R. Hefner, D. Y. Chen, and F. C. Lee, Insulated Gate Bipolar Transistor (IGBT) Modelling Using IG-SPICE, *EEE Trans. Ind. Appl.*, Vol. 30, No.1, January/February 1994, pp. 24–33.

29. A. R. Hefner, Jr., and D. M. Diebolt, An Experimentally Verified IGBT Model Implemented in the Saber Circuit Simulator, *IEEE Trans. Power Electron.*, Vol. 9, No. 5, September 1994, pp. 532–542.
30. A. N. Githiari, B. M. Gordon, R. A. McMahon, Z.-M. Li, and P. A. Mawby, A Comparison of IGBT Models for Use in Circuit Design, *IEEE Trans. Power Electron.*, Vol. 4, No. 4, July 1999, pp. 607–614.

APPENDIX 5.1: SOLUTION OF EQ. (5.8)

The solution of the differential equation

$$\frac{d^2y}{dx^2} + a\frac{dy}{dx} + by = f(x) \tag{A5.1.1}$$

is the sum of two components:

(i) The *complementary function* (C.F.) given by

$$\text{C.F.} = A \exp \lambda_1 x + B \exp \lambda_2 x \tag{A5.1.2}$$

where A, B are constants and λ_1, λ_2 are the roots of the equation

$$\lambda^2 + a\lambda + b = 0 \tag{A5.1.3}$$

(ii) The particular integral

By analogy, the solution of Eq. (5.8) is

$$p(x) = A \exp\left(+ \frac{d-x}{L_a} \right) + B \exp\left(- \frac{d-x}{L_a} \right) \tag{A5.1.5}$$

where the high-level diffusion length is taken as the ambipolar diffusion length, that is, $L_{HL} = L_a$. This solution is obtained since

$$\lambda^2 - \frac{1}{L_{HL}{}^2} = 0 \tag{A5.1.6}$$

giving

$$\lambda = \pm \frac{1}{L_{HL}} \tag{A5.1.7}$$

and the particular integral is obviously zero.

Applying the boundary condition, for $x = 0$, $p(x) = p_0$, we have

$$p_0 = A \exp(d/L_a) + B \exp(-d/L_a) \tag{A5.1.8}$$

Also, at $x = d$, $p(d) = 0$. Hence,

$$A + B = 0 \tag{A5.1.9}$$

In Eq. (A5.1.8), putting $B = -A$ from Eq. (A5.1.9) and expanding exponentials in terms of hyperbolic cosine and hyperbolic sine as $\exp(z) = \cosh z + \sinh z$, and $\exp(-z) = \cosh z - \sinh z$, we get

$$\begin{aligned} p_0 &= A\{\cosh(d/L_a) + \sinh(d/L_a)\} + B\{\cosh(d/L_a) - \sinh(d/L_a)\} \\ &= 2A\sinh(d/L_a) \end{aligned} \tag{A5.1.10}$$

which yields

$$A = \frac{p_0}{2\sinh(d/L_a)} \tag{A5.1.11}$$

Putting the value of A in Eq. (A5.1.9), we have

$$B = -\frac{p_0}{2\sinh(d/L_a)} \tag{A.5.1.12}$$

Then substituting the values of A and B from Eqs. (A.5.1.11) and (A.5.1.12) into Eq. (A5.1.5) and expanding the exponentials in Eq. (A5.1.5) in hyperbolic cosines and hyperbolic sines, we arrive at

$$\begin{aligned} p(x) &= \frac{p_0}{2\sinh(d/L_a)}\left\{\cosh\left(\frac{d-x}{L_a}\right) + \sinh\left(\frac{d-x}{L_a}\right)\right\} \\ &\quad - \frac{p_0}{2\sinh(d/L_a)}\left\{\cosh\left(\frac{d-x}{L_a}\right) - \sinh\left(\frac{d-x}{L_a}\right)\right\} \\ &= p_0\frac{\sinh\{(d-x)/L_a\}}{\sinh(d/L_a)} \end{aligned} \tag{A5.1.13}$$

which is Eq. (5.9).

APPENDIX 5.2: DERIVATION OF EQS. (5.33) AND (5.34)

The continuity equations for electrons and holes are written as [1]

$$\frac{\partial n}{\partial t} = \frac{1}{q}\frac{\partial J_n}{\partial t} - \left(\frac{n}{\tau_n} - G_0\right) \tag{A5.2.1}$$

$$\frac{\partial p}{\partial t} = -\frac{1}{q}\frac{\partial J_p}{\partial t} - \left(\frac{p}{\tau_p} - G_0\right) \tag{A5.2.2}$$

where n and p are the excess electron and hole concentrations, τ_n and τ_n are the mean lifetimes of electrons and holes, respectively, and G_0 is the thermal generation rate. The current densities J_n and J_p of the electrons and holes in an electric field E are expressed as

$$J_n = \sigma_n E + qD_n \frac{\partial n}{\partial x} \tag{A5.2.3}$$

$$J_p = \sigma_p E - qD_p \frac{\partial p}{\partial x} \tag{A5.2.4}$$

The total current density is given by

$$J_{\text{CEI}} = J_n + J_p \tag{A5.2.5}$$

It is solenoidal so that

$$\frac{\partial J}{\partial x} = 0 \tag{A5.2.6}$$

σ_n and σ_p denote the contributions from electrons and holes to the total conductivity σ:

$$\sigma_n = q\mu_n n, \qquad \sigma_p = q\mu_p p, \qquad \sigma = \sigma_n + \sigma_p \tag{A5.2.7}$$

D_n and D_p are the diffusion coefficients of electrons and holes, respectively. For the ambipolar transport of added carriers n or p, the condition of neutrality requires that

$$n = p \tag{A5.2.8}$$

Since $J = \sigma E$, using Eqs. (A5.2.5) and (A5.2.7), the electric field E is eliminated from Eqs. (A5.2.3) and (A5.2.4). In terms of the electron and hole currents I_n, I_p and the total current I_{CEI}, Eqs. (A5.2.3) and (A5.2.4) are expressed as

$$I_n = \frac{\sigma_n}{\sigma} I_{\text{CEI}} + qAD \frac{\partial n}{\partial x} \tag{A5.2.9}$$

$$I_p = \frac{\sigma_p}{\sigma} I_{\text{CEI}} - qAD \frac{\partial p}{\partial x} \tag{A5.2.10}$$

where A is the device area and a general concentration diffusion coefficient D_a for the carriers is introduced as

$$D_a \equiv \frac{\sigma_n D_p + \sigma_p D_n}{\sigma} = \frac{n + p}{n/D_p + p/D_n} \tag{A5.2.11}$$

It is called the *ambipolar diffusion coefficient.* From Eqs. (A5.2.7), we have

$$\frac{\sigma_{\mathrm{n}}}{\sigma}=\frac{1}{1+(1/b)(p/n)} \tag{A5.2.12}$$

$$\frac{\sigma_{\mathrm{p}}}{\sigma}=\frac{1}{1+b(n/p)} \tag{A5.2.13}$$

where b is the ambipolar mobility ratio $=\mu_{\mathrm{n}}/\mu_{\mathrm{p}}$. By Eq. (A5.2.8), Eqs. (A5.2.12) and (A5.2.13) reduce to the form

$$\frac{\sigma_{\mathrm{n}}}{\sigma}=\frac{1}{1+(1/b)}=\frac{b}{1+b} \tag{A5.2.14}$$

$$\frac{\sigma_{\mathrm{p}}}{\sigma}=\frac{1}{1+b} \tag{A5.2.15}$$

Substituting for $\sigma_{\mathrm{n}}/\sigma$ and $\sigma_{\mathrm{p}}/\sigma$ in Eqs. (A5.2.9) and (A5.2.10), we get

$$I_{\mathrm{n}}=\frac{b}{1+b}I_{\mathrm{CEI}}+qAD\frac{\partial n}{\partial x} \tag{A5.2.16}$$

$$I_{\mathrm{p}}=\frac{1}{1+b}I_{\mathrm{CEI}}-qAD\frac{\partial p}{\partial x} \tag{A5.2.17}$$

REFERENCES

1. W. Van Roosbroeck, The Transport of Added Carriers in a Homogeneous Semiconductor, *Phys. Rev.*, Vol. 91, No. 2, July 15, 1953, pp. 282–289.

APPENDIX 5.3: DERIVATION OF EQ. (5.35)

In a semiconductor, the derivation of minority-carrier current and density is based on the assumptions of (i) the prevalence of space-charge neutrality condition and (ii) the minority carriers as the only type of carriers present. These minority carriers are transported by the diffusion mechanism. This is known as the *diffusion approximation.* The hole current in N-type semiconductor is obtained by setting the electric field $\xi=0$ in the transport equation

$$I_{\mathrm{p}}=qA\mu_{\mathrm{p}}\Delta p\xi-qAD_{\mathrm{p}}\frac{d\Delta p}{dx}=-qAD_{\mathrm{p}}\frac{d\Delta p}{dx} \tag{A5.3.1}$$

where q is the electronic charge, A is the area, Δp is the excess hole concentration, and D_{p} is the diffusion coefficient of holes. Then the hole continuity equation

$$\frac{1}{qA}\frac{\partial I_p}{\partial x} + \frac{\Delta p}{\tau_p} = -\frac{\partial \Delta p}{\partial t} \tag{A5.3.2}$$

becomes

$$-D_p\frac{\partial^2 \Delta p}{\partial x^2} + \frac{\Delta p}{\tau_p} = -\frac{\partial \Delta p}{\partial t} \tag{A5.3.3}$$

which may be recast as

$$\frac{\partial^2 \Delta p}{\partial x^2} = \frac{\Delta p}{D_p \tau_p} + \frac{1}{D_p}\frac{\partial \Delta p}{\partial t} \tag{A5.3.4}$$

If L_p denotes the diffusion length of holes, then $L_p = \sqrt{(D_p \tau_p)}$ and Eq. (A5.3.4) becomes

$$\frac{\partial^2 \Delta p}{\partial x^2} = \frac{\Delta p}{L_p^2} + \frac{1}{D_p}\frac{\partial \Delta p}{\partial t} \tag{A5.3.5}$$

Under high-level injection conditions, the effect of both electrons and holes must be taken into account to calculate the current flow, and the minority-carrier (or hole) transport is expressed in the form

$$\frac{\partial^2 \Delta p}{\partial x^2} = \frac{\Delta p}{L_a^2} + \frac{1}{D_a}\frac{\partial \Delta p}{\partial t} \tag{A.3.6}$$

where D_a and L_a are defined as the ambipolar diffusion coefficient and ambipolar diffusion length, respectively. D_a is given by

$$D_a = \frac{(n+p)D_n D_p}{nD_n + pD_p} \tag{A5.3.7}$$

where n and p are the electron and hole concentrations and D_n and D_p the corresponding diffusion coefficients.

APPENDIX 5.4: DERIVATION OF EQ. (5.38) [SOLUTION OF EQ. (5.35)]

The solution to the steady-state ambipolar diffusion equation (5.35)

$$\frac{\partial^2 \Delta p(x)}{\partial x^2} = \frac{\Delta p(x)}{L_a^2} \tag{A5.4.1}$$

with $(\partial/\partial t)\Delta p(x) = 0$ is written as

$$\Delta p(x) = A\cosh(x/L_a) + B\sinh(x/L_a) \tag{A5.4.2}$$

where the constants of integration A and B are evaluated from the boundary conditions

$$\Delta p(x)|_{x=0} = p_0 \tag{A5.4.3}$$

$$\Delta p(x)|_{x=W} = 0 \tag{A5.4.4}$$

Applying the boundary condition (A5.4.3) to Eq. (5.4.2), we get

$$p_0 = A \tag{A5.4.5}$$

By applying the boundary condition (A5.4.4) to Eq. (A5.4.2), we have

$$0 = A\cosh(W/L_a) + B\sinh(W/L_a) \tag{A5.4.6}$$

giving

$$B = -\frac{p_0\cosh(W/L_a)}{\sinh(W/L_a)} \tag{A5.4.7}$$

Substituting the values of A and B from Eqs. (A5.4.5) and (A5.4.7) into Eq. (A5.4.2), we obtain

$$\Delta p(x) = p_0\cosh(x/L_a) - \frac{p_0\cosh(W/L_a)}{\sinh(W/L_a)}\sinh(x/L_a)$$

$$= \frac{p_0\{\sinh(W/L_a)\cosh(x/L_a) - \cosh(W/L_a)\sinh(x/L_a)\}}{\sinh(W/L_a)}$$

or

$$\Delta p(x) = p_0\frac{\sinh\left(\dfrac{W-x}{L_a}\right)}{\sinh(W/L_a)} \tag{A5.4.8}$$

APPENDIX 5.5: DERIVATION OF EQS. (5.40)–(5.42)

Let us direct our attention to the forward-biased P$^+$-emitter/N$^-$-base diode of the PNP transistor in the IGBT. The quasi-equilibrium approximation is assumed to be valid. This implies that the product pn = constant throughout the space-charge region of the diode. The expression for the electron current injected into the emitter is

$$I_n|_{x=0} = \frac{qAD_n n_{p0}}{L_n}\left\{\exp\left(\frac{qV}{kT}\right) - 1\right\} \cong I_{sn}\left\{\exp\left(\frac{qV}{kT}\right)\right\} \qquad \text{(A5.5.1)}$$

where q is the electronic charge, A is the junction area, D_n is the electron diffusion coefficients, L_n is the diffusion length of electrons, and n_{p0} is the equilibrium electron density on the P side, V is the applied forward voltage, k is the Boltzmann's constant, and T is the absolute temperature. The quantity $qAD_n n_{p0}/L_n$ is the emitter electron saturation current.

But the excess hole concentration at the edge of the emitter–base space charge region ($x = 0$) is given by

$$p_0 = p_{n0}\exp\left(\frac{qV}{kT}\right) = \frac{n_i^2}{N_D}\exp\left(\frac{qV}{kT}\right) \qquad \text{(A5.5.2)}$$

where p_{n0} is the equilibrium hole concentration on the N side $= n_i^2/N_D$ with n_i = intrinsic carrier concentration and N_D = donor concentration of N side. This equation gives

$$\exp\left(\frac{qV}{kT}\right) = \frac{p_0 N_D}{n_i^2} \qquad \text{(A5.5.3)}$$

Under high-level injection, the carrier concentration in the N^- base equals $p_0 + N_D$, where $p_0 > N_D$. Then Eq. (A5.5.3) is modified to

$$\exp\left(\frac{qV}{kT}\right) = \frac{p_0(p_0 + N_D)}{n_i^2} \qquad \text{(A5.5.4)}$$

Neglecting N_D, Eq. (A5.5.4) simplifies to

$$\exp\left(\frac{qV}{kT}\right) = \frac{p_0^2}{n_i^2} \qquad \text{(A5.5.5)}$$

Using Eqs. (A5.5.1) and (A5.5.5), we obtain

$$I_n|_{x=0} = I_{sn}\frac{p_0^2}{n_i^2} \qquad \text{(A5.5.6)}$$

Differentiating Eq. (5.38) with respect to x, we get

$$\frac{d\Delta p(x)}{dx} = -\frac{p_0}{L_\mathrm{a}} \frac{\cosh\left(\dfrac{W-x}{L_\mathrm{a}}\right)}{\sinh(W/L_\mathrm{a})} \qquad \text{(A5.5.7)}$$

Substituting for $d\Delta p/dx$ in Eqs. (5.33) and (5.34), we have

$$I_\mathrm{n}(x) = \frac{b}{1+b} I_\mathrm{CEI} - \frac{qAp_0D_\mathrm{a}}{L_\mathrm{a}} \frac{\cosh\left(\dfrac{W-x}{L_\mathrm{a}}\right)}{\sinh(W/L_\mathrm{a})} \qquad \text{(A5.5.8)}$$

$$I_\mathrm{p}(x) = \frac{1}{1+b} I_\mathrm{CEI} + \frac{qAp_0D_\mathrm{a}}{L_\mathrm{a}} \frac{\cosh\left(\dfrac{W-x}{L_\mathrm{a}}\right)}{\sinh(W/L_\mathrm{a})} \qquad \text{(A5.5.9)}$$

Putting successively $x = W$, 0 in Eq. (A5.5.8), we obtain

$$I_\mathrm{n}(W) = \frac{b}{1+b} I_\mathrm{CEI} - \frac{qAp_0D_\mathrm{a}}{L_\mathrm{a}} \frac{1}{\sinh(W/L_\mathrm{a})} \qquad \text{(A5.5.10)}$$

$$I_\mathrm{n}(0) = \frac{b}{1+b} I_\mathrm{CEI} - \frac{qAp_0D_\mathrm{a}}{L_\mathrm{a}} \coth(W/L_\mathrm{a}) \qquad \text{(A5.5.11)}$$

Subtracting Eq. (A5.5.11) from Eq. (A5.5.10), we get

$$I_\mathrm{n}(W) - I_\mathrm{n}(0) = \frac{qAp_0D_\mathrm{a}}{L_\mathrm{a}} \left\{ \coth(W/L_\mathrm{a}) - \frac{1}{\sinh(W/L_\mathrm{a})} \right\} \qquad \text{(A5.5.12)}$$

Substituting for $I_\mathrm{n}(0)$ from Eq. (5.40), we have

$$I_\mathrm{n}(W) = \frac{{p_0}^2 I_\mathrm{s}}{n_\mathrm{i}^2} + \frac{qAp_0D_\mathrm{a}}{L_\mathrm{a}} \left\{ \coth(W/L_\mathrm{a}) - \frac{1}{\sinh(W/L_\mathrm{a})} \right\} \qquad \text{(A5.5.13)}$$

Now, putting $x = W$ in Eq. (5.5.9), we obtain

$$I_\mathrm{p}(W) = \frac{1}{1+b} I_\mathrm{CEI} + \frac{qAp_0D_\mathrm{a}}{L_\mathrm{a}} \frac{1}{\sinh(W/L_\mathrm{a})} \qquad \text{(A5.5.14)}$$

Also, dividing both sides of Eq. (A5.5.11) by b, we get

$$\frac{I_{\mathrm{n}}(0)}{b} = \frac{1}{1+b} I_{\mathrm{CEI}} - \frac{qAp_0 D_{\mathrm{a}}}{L_{\mathrm{a}}} \frac{\coth(W/L_{\mathrm{a}})}{b} \tag{A5.5.15}$$

Subtracting Eq. (A5.5.15) from Eq. (A5.5.14), we have

$$I_{\mathrm{p}}(W) - \frac{I_{\mathrm{n}}(0)}{b} = \frac{qAp_0 D_{\mathrm{a}}}{L_{\mathrm{a}}} \left\{ \frac{\coth(W/L_{\mathrm{a}})}{b} - \frac{1}{\sinh(W/L_{\mathrm{a}})} \right\} \tag{A5.5.16}$$

or

$$I_{\mathrm{p}}(W) = \frac{p_0 I_{\mathrm{s}}}{b n_{\mathrm{i}}^2} + \frac{qAp_0 D_{\mathrm{a}}}{L_{\mathrm{a}}} \left\{ \frac{\coth(W/L_{\mathrm{a}})}{b} - \frac{1}{\sinh(W/L_{\mathrm{a}})} \right\} \tag{A5.5.17}$$

APPENDIX 5.6: DERIVATION OF EQ. (5.44)

The potential drop $V_{\mathrm{N^-}}$ across the N-base region is determined by integrating the equation

$$\frac{dV_{\mathrm{N^-}}}{dx} = -\frac{I_{\mathrm{n}}(x)}{qA\mu_{\mathrm{n}} n(x)} \tag{A5.6.1}$$

from $x = 0$ to $x = W$ with I_{n} given by Eq. (5.33) and $n(x) = N_{\mathrm{D}} + \Delta p(x)$, where $\Delta p(x)$ is expressed by Eq. (5.38). Considering the first term of Eq. (5.33), we have

$$\begin{aligned} V_{\mathrm{N_-}} &= -\int_0^W \frac{b}{1+b} \frac{I_{\mathrm{CEI}}(x)\,dx}{qA\mu_{\mathrm{n}} n(x)} = -\frac{I_{\mathrm{CEI}}(x)}{(b+1/b)} \frac{1}{\mu_{\mathrm{n}} Aq} \int_0^W \frac{dn}{n(x)} \\ &= -\frac{I_{\mathrm{CEI}}(x)}{(b+1/b)} \frac{1}{\mu_{\mathrm{n}} Aq} \int_0^W \frac{dx}{\left\{ N_{\mathrm{D}} + p_0 \sinh\left(\dfrac{W-x}{L_{\mathrm{a}}} \right) \operatorname{cosech}(W/L_{\mathrm{a}}) \right\}} \end{aligned} \tag{A5.6.2}$$

This integral is converted into the form $\int dz/(a^2 - z^2) = \{1/(2a)\} \ln\{(a+z)/(a-z)\}$ + integration constant, by a transformation of variables accomplished by putting $\tanh\{(W-x)/2L_{\mathrm{a}})\} = t$. Then $\operatorname{cosech}^2\{(W-x)/2L_{\mathrm{a}})\} = \coth^2\{(W-x)/2L_{\mathrm{a}})\} - 1 = 1/\tanh^2\{(W-x)/2L_{\mathrm{a}})\} - 1 = 1/t^2 - 1 = (1-t^2)/t^2$. $\therefore \sinh\{(W-x)/2L_{\mathrm{a}})\} = t/\sqrt{1-t^2}$. Also, $\cosh^2\{(W-x)/2L_{\mathrm{a}})\} - \sinh^2\{(W-x)/2L_{\mathrm{a}})\} = 1$. $\therefore \cosh^2\{(W-x)/2L_{\mathrm{a}})\} = 1 + \sinh^2\{(W-x)/$

$2L_a)\} = 1 + t^2/(1 - t^2) = (1 - t^2 + t^2)/(1 - t^2) = 1/(1 - t^2)$. $\therefore \cosh\{(W - x)/L_a)\} = 1/\sqrt{1-t^2}$. $\therefore \sinh\{(W - x)/L_a)\} = 2\sinh\{(W - x)/2L_a)\} \times \cosh\{(W - x)/2L_a)\} = 2\{t/\sqrt{1-t^2}\} \times 1/\sqrt{1-t^2} = 2t/(1 - t^2)$. Also, $dt/dx = (d/dx)[\tanh\{(W - x)/2L_a)\} = 1/(2L_a)\mathrm{sech}^2\{(W - x)/2L_a)\} = 1/(2L_a)[1 - \tanh^2\{(W - x)/2L_a)\} = \{1/(2L_a)\}(1 - t^2)$. $\therefore dx/dt = 2L_a/(1 - t^2)$, and $dx = 2L_a dt/(1 - t^2)$. Hence, Eq. (5.6.2) becomes

$$V_{\mathrm{N-}} = -\frac{I_{\mathrm{CEI}}(x)}{(b+1/b)}\frac{2L_a}{\mu_n A q}\int_0^W \frac{\dfrac{dt}{1-t^2}}{\left\{N_D + p_0\dfrac{2t}{1-t^2}\,\mathrm{cosech}\left(\dfrac{W}{L_a}\right)\right\}}$$

$$= -\frac{I_{\mathrm{CEI}}(x)}{(b+1/b)}\frac{2L_a}{\mu_n A q}\int_0^W \frac{dt}{\left\{N_D(1-t^2) + 2p_0 t \times \mathrm{cosech}\left(\dfrac{W}{L_a}\right)\right\}}$$

$$= -\frac{I_{\mathrm{CEI}}(x)}{(b+1/b)}\frac{2L_a}{\mu_n A q N_D}\int_0^W \frac{dt}{\left\{(1-t^2) + \dfrac{2p_0 t}{N_D} \times \mathrm{cosech}\left(\dfrac{W}{L_a}\right)\right\}}$$

$$= -\frac{I_{\mathrm{CEI}}(x)}{(b+1/b)}\frac{2L_a}{\mu_n A q N_D}\int_0^W \frac{dt}{-\left\{\begin{array}{l} t^2 - 2t\times\dfrac{p_0}{N_D}\,\mathrm{cosech}\left(\dfrac{W}{L_a}\right) \\ +\dfrac{p_0^2}{N_D{}^2}\,\mathrm{cosech}^2\left(\dfrac{W}{L_a}\right)\end{array}\right\} + 1 + \dfrac{p_0{}^2}{N_D{}^2}\,\mathrm{cosech}^2\left(\dfrac{W}{L_a}\right)}$$

$$= -\frac{I_{\mathrm{CEI}}(x)}{(b+1/b)}\frac{2L_a}{\mu_n A q N_D}\int_0^W \frac{dt}{\left\{\sqrt{1+\dfrac{p_0{}^2}{N_D{}^2}\,\mathrm{cosech}^2\left(\dfrac{W}{L_a}\right)}\right\}^2 - \left\{t - \dfrac{p_0}{N_D}\,\mathrm{cosech}\left(\dfrac{W}{L_a}\right)\right\}^2}$$

$$= -\frac{I_{\mathrm{CEI}}(x)}{(b+1/b)}\frac{2L_a}{\mu_n A q N_D} \times \frac{1}{2\left\{\sqrt{1+\dfrac{p_0{}^2}{N_D{}^2}\,\mathrm{cosech}^2\left(\dfrac{W}{L_a}\right)}\right\}}$$

$$\times \ln\left[\frac{\left\{\sqrt{1+\dfrac{p_0{}^2}{N_D{}^2}\,\mathrm{cosech}^2\left(\dfrac{W}{L_a}\right)}\right\} + \left\{t - \dfrac{p_0}{N_D}\,\mathrm{cosech}\left(\dfrac{W}{L_a}\right)\right\}}{\left\{\sqrt{1+\dfrac{p_0{}^2}{N_D{}^2}\,\mathrm{cosech}^2\left(\dfrac{W}{L_a}\right)}\right\} - \left\{t - \dfrac{p_0}{N_D}\,\mathrm{cosech}\left(\dfrac{W}{L_a}\right)\right\}}\right]_0^W$$

$$= -\frac{I_{\text{CEI}}(x)}{(b+1/b)}\frac{2L_a}{\mu_n A q N_D}\times\frac{N_D}{2\left\{\sqrt{N_D{}^2+p_0{}^2\operatorname{cosech}^2\left(\frac{W}{L_a}\right)}\right\}}$$

$$\times\ln\left[\frac{\dfrac{\left\{\sqrt{N_D{}^2+p_0{}^2\operatorname{cosech}^2\left(\frac{W}{L_a}\right)}\right\}+\left\{N_D t-p_0\operatorname{cosech}\left(\frac{W}{L_a}\right)\right\}}{N_D}}{\dfrac{\left\{\sqrt{N_D{}^2+p_0{}^2\operatorname{cosech}^2\left(\frac{W}{L_a}\right)}\right\}-\left\{N_D t-p_0\operatorname{cosech}\left(\frac{W}{L_a}\right)\right\}}{N_D}}\right]_0^W$$

$$= -\frac{I_{\text{CEI}}(x)}{(b+1/b)}\frac{L_a}{\mu_n A q}\times\frac{1}{\left\{\sqrt{N_D{}^2+p_0{}^2\operatorname{cosech}^2\left(\frac{W}{L_a}\right)}\right\}}$$

$$\times\ln\left[\frac{\left\{\sqrt{N_D{}^2+p_0{}^2\operatorname{cosech}^2\left(\frac{W}{L_a}\right)}\right\}+\left\{N_D t-p_0\operatorname{cosech}\left(\frac{W}{L_a}\right)\right\}}{\left\{\sqrt{N_D{}^2+p_0{}^2\operatorname{cosech}^2\left(\frac{W}{L_a}\right)}\right\}-\left\{N_D t-p_0\operatorname{cosech}\left(\frac{W}{L_a}\right)\right\}}\right]_0^W$$

$$= -\frac{I_{\text{CEI}}(x)}{(b+1/b)}\frac{L_a}{\mu_n A q}\times\frac{1}{\left\{\sqrt{N_D{}^2+p_0{}^2\operatorname{cosech}^2\left(\frac{W}{L_a}\right)}\right\}}$$

$$\times\ln\left[\frac{1+\dfrac{\left\{N_D t-p_0\operatorname{cosech}\left(\frac{W}{L_a}\right)\right\}}{\left\{\sqrt{N_D{}^2+p_0{}^2\operatorname{cosech}^2\left(\frac{W}{L_a}\right)}\right\}}}{1-\dfrac{\left\{N_D t-p_0\operatorname{cosech}\left(\frac{W}{L_a}\right)\right\}}{\left\{\sqrt{N_D{}^2+p_0{}^2\operatorname{cosech}^2\left(\frac{W}{L_a}\right)}\right\}}}\right]_0^W$$

$$= -\frac{I_{\text{CEI}}(x)}{(b+1/b)}\frac{L_a}{\mu_n A q}\times\frac{1}{\left\{\sqrt{N_D{}^2+p_0{}^2\operatorname{cosech}^2\left(\frac{W}{L_a}\right)}\right\}}$$

$$\times\left[2\,\mathrm{arctanh}^{-1}\frac{\left\{N_{\mathrm{D}}t-p_0\,\mathrm{cosech}\left(\dfrac{W}{L_{\mathrm{a}}}\right)\right\}}{\left\{\sqrt{N_{\mathrm{D}}{}^2+p_0{}^2\,\mathrm{cosech}^2\left(\dfrac{W}{L_{\mathrm{a}}}\right)}\right\}}\right]_0^W$$

$$=-\frac{I_{\mathrm{CEI}}(x)}{(b+1/b)}\frac{L_{\mathrm{a}}}{\mu_{\mathrm{n}}Aq}\times\frac{2}{\left\{\sqrt{N_{\mathrm{D}}{}^2+p_0{}^2\,\mathrm{cosech}^2\left(\dfrac{W}{L_{\mathrm{a}}}\right)}\right\}}$$

$$\times\left[\mathrm{arctanh}^{-1}\frac{\left\{N_{\mathrm{D}}\tanh\left(\dfrac{W-x}{2L_{\mathrm{a}}}\right)-p_0\,\mathrm{cosech}\left(\dfrac{W}{L_{\mathrm{a}}}\right)\right\}}{\left\{\sqrt{N_{\mathrm{D}}{}^2+p_0{}^2\,\mathrm{cosech}^2\left(\dfrac{W}{L_{\mathrm{a}}}\right)}\right\}}\right]_0^W \qquad (\mathrm{A5.6.3})$$

where the formula $(1/2)\ln\{(1+x)/(1-x)\}=\tanh^{-1}x$ or, $\ln\{(1+x)/(1-x)\}=2\tanh^{-1}x$ has been applied. Because $\tanh^{-1}A-\tanh^{-1}B=\tanh^{-1}\{(A+B)/(1-AB)\}$, the square bracketed term in Eq. (A5.6.3) simplifies to

$$=\mathrm{arctanh}^{-1}\frac{\left\{-p_0\,\mathrm{cosech}\left(\dfrac{W}{L_{\mathrm{a}}}\right)\right\}}{\left\{\sqrt{N_{\mathrm{D}}{}^2+p_0{}^2\mathrm{cosech}^2\left(\dfrac{W}{L_{\mathrm{a}}}\right)}\right\}}$$

$$-\,\mathrm{arctanh}^{-1}\frac{\left\{N_{\mathrm{D}}\tanh\left(\dfrac{W}{2L_{\mathrm{a}}}\right)-p_0\,\mathrm{cosech}\left(\dfrac{W}{L_{\mathrm{a}}}\right)\right\}}{\left\{\sqrt{N_{\mathrm{D}}{}^2+p_0{}^2\mathrm{cosech}^2\left(\dfrac{W}{L_{\mathrm{a}}}\right)}\right\}}$$

$$=\mathrm{arctanh}^{-1}\frac{\dfrac{\left\{-p_0\,\mathrm{cosech}\left(\dfrac{W}{L_{\mathrm{a}}}\right)\right\}-\left\{N_{\mathrm{D}}\tanh\left(\dfrac{W}{2L_{\mathrm{a}}}\right)-p_0\,\mathrm{cosech}\left(\dfrac{W}{L_{\mathrm{a}}}\right)\right\}}{\left\{\sqrt{N_{\mathrm{D}}{}^2+p_0{}^2\mathrm{cosech}^2\left(\dfrac{W}{L_{\mathrm{a}}}\right)}\right\}}}{1-\dfrac{\left\{-p_0\,\mathrm{cosech}\left(\dfrac{W}{L_{\mathrm{a}}}\right)\right\}\times\left\{N_{\mathrm{D}}\tanh\left(\dfrac{W}{2L_{\mathrm{a}}}\right)-p_0\,\mathrm{cosech}\left(\dfrac{W}{L_{\mathrm{a}}}\right)\right\}}{N_{\mathrm{D}}{}^2+p_0{}^2\mathrm{cosech}^2\left(\dfrac{W}{L_{\mathrm{a}}}\right)}}$$

$$
\begin{aligned}
&= \operatorname{arctanh}^{-1} \frac{\dfrac{\left\{ \begin{array}{l} \left\{-p_0 \operatorname{cosech}\left(\dfrac{W}{L_a}\right)\right\} - \left\{N_D \tanh\left(\dfrac{W}{2L_a}\right) - p_0 \operatorname{cosech}\left(\dfrac{W}{L_a}\right)\right\} \\ \times \left\{N_D{}^2 + p_0{}^2 \operatorname{cosech}^2\left(\dfrac{W}{L_a}\right)\right\} \end{array} \right\}}{\left\{\sqrt{N_D{}^2 + p_0{}^2 \operatorname{cosech}^2\left(\dfrac{W}{L_a}\right)}\right\}}}{\left\{ \begin{array}{l} N_D{}^2 + p_0{}^2 \operatorname{cosech}^2\left(\dfrac{W}{L_a}\right) - \left\{-p_o \operatorname{cosech}\left(\dfrac{W}{L_a}\right)\right\} \\ \times \left\{N_D \tanh\left(\dfrac{W}{2L_a}\right) - p_0 \operatorname{cosech}\left(\dfrac{W}{L_a}\right)\right\} \end{array} \right\}} \\[2ex]
&= \operatorname{arctanh}^{-1} \frac{\left\{ \begin{array}{l} \left[\left\{-p_0 \operatorname{cosech}\left(\dfrac{W}{L_a}\right)\right\} - \left\{N_D \tanh\left(\dfrac{W}{2L_a}\right) - p_0 \operatorname{cosech}\left(\dfrac{W}{L_a}\right)\right\}\right] \\ \times \left\{\sqrt{N_D{}^2 + p_0{}^2 \operatorname{cosech}^2\left(\dfrac{W}{L_a}\right)}\right\} \end{array} \right\}}{\left\{ \begin{array}{l} N_D{}^2 + p_0{}^2 \operatorname{cosech}^2\left(\dfrac{W}{L_a}\right) - \left\{-p_o \operatorname{cosech}\left(\dfrac{W}{L_a}\right)\right\} \\ \times \left\{N_D \tanh\left(\dfrac{W}{2L_a}\right) - p_0 \operatorname{cosech}\left(\dfrac{W}{L_a}\right)\right\} \end{array} \right\}} \\[2ex]
&= \operatorname{arctanh}^{-1} \frac{-N_D \tanh\left(\dfrac{W}{L_a}\right) \times \left\{\sqrt{N_D{}^2 + p_0{}^2 \operatorname{cosech}^2\left(\dfrac{W}{L_a}\right)}\right\}}{N_D{}^2 - \left\{-p_0 \operatorname{cosech}\left(\dfrac{W}{L_a}\right)\right\} N_D \tanh\left(\dfrac{W}{2L_a}\right)} \\[2ex]
&= -\operatorname{arctanh}^{-1} \frac{\tanh\left(\dfrac{W}{2L_a}\right) \times \left\{\sqrt{N_D{}^2 + p_0{}^2 \operatorname{cosech}^2\left(\dfrac{W}{L_a}\right)}\right\}}{N_D + p_0 \operatorname{cosech}\left(\dfrac{W}{L_a}\right) \tanh\left(\dfrac{W}{2L_a}\right)} \qquad \text{(A5.6.4)}
\end{aligned}
$$

because $\tanh^{-1}(x) = -\tanh^{-1}(x)$.

APPENDIX 5.7: DERIVATION OF EQS. (5.81) AND CONSTRUCTION OF EQUIVALENT CONDUCTIVE NETWORK FOR 1-D LINEAR ELEMENT

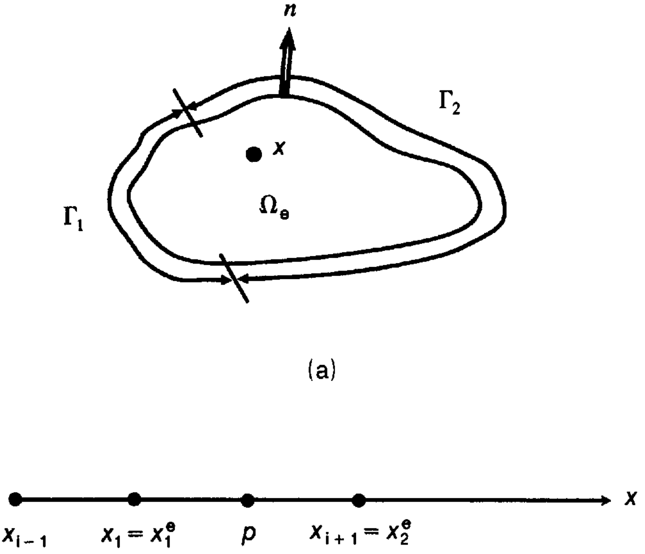
Considering the semidiscrete heat equation (5.80), the temperature $T(x, t) = \Sigma_i^n d_i(t) N_i(x)$, where d_i is the temperature at node i and N_i is the corresponding finite element basis function. Assembly of the global matrices **M**, **K**, and $\mathbf{F}$ is carried out from the elemental matrices $\mathbf{m}_e$, $\mathbf{k}_e$, and $\mathbf{f}_e$, expressed as [1]

$$\mathbf{M} = \bigoplus_{e=1}^{ncl} \mathbf{m}_e \quad \text{with} \quad \mathbf{m}_e = [(m_e)_{ij}], \quad \text{where } (m_e)_{ij} = \rho c_p A \int_{\Omega_e} N_i N_j \, d\Omega_e \qquad \text{(A5.7.1)}$$

$$\mathbf{K} = \bigoplus_{e=1}^{ncl} \mathbf{k}_e \quad \text{with} \quad \mathbf{k}_e = [(k_e)_{ij}], \quad \text{where } (m_e)_{ij} = kA \int_{\Omega_e} \nabla N_i \cdot \nabla N_j \, d\Omega_e + hA \int_{\Gamma_2} N_i N_j \, d\Gamma_2 \qquad \text{(A5.7.2)}$$

$$\mathbf{F} = \bigoplus_{e=1}^{ncl} \mathbf{f}_e \quad \text{with} \quad \mathbf{f}_e = [(f_e)_{ij}], \quad \text{where } (m_e)_{ij} = P \int_{\Gamma_1} N_i \, d\Gamma_1 + hAT_a \int_{\Gamma_2} N_i \, d\Gamma_2 + A \int_{\Omega_e} g N_i \, d\Omega_e \qquad \text{(A5.7.3)}$$

where Ω_e, Γ_1, and Γ_2 are the boundaries defining the semiconductor self-heating, the electrical and the thermal systems respectively of the mathematical domain in the problem (Fig. A5.7a), ρ is the mass density, c_p is the specific heat, A is the cross-sectional area of the input power, k is the



(a)

(b)

Figure A5.7.1 (a) Mathematical domain of the problem, and (b) length coordinates.

thermal conductivity, h is the convection coefficient, P is the input power from boundary Γ_1, g is the generating heat, and T_a is the ambient temperature. In these equations, Eq. (A5.7.1) and the first term of Eq. (A5.7.2) express conduction phenomena, the second terms of Eqs. (A5.7.2) and (5.7.3) represent convective effects, and the first and the third terms of Eq. (A5.7.3) refer to input power and heat generation effects respectively.

For deriving the one-dimensional equivalent thermal network, the following length coordinates are used (Fig. A5.7b):

$$L_1 = \frac{\text{Distance }(px_2^e)}{\text{Distance }(x_1^e x_2^e)} = \frac{x_2^e - x}{\delta_x}, \quad L_2 = \frac{\text{Distance }(px_1^e)}{\text{Distance }(x_1^e x_2^e)} = \frac{x - x_1^e}{\delta_x} \qquad \text{(A5.7.4)}$$

where $\delta_x = x_2^e - x_1^e$, $(L_1, L_2) = (1, 0)$ at $x = x_1^e$, and $(L_1, L_2) = (0, 1)$ at $x = x_2^e$. Calculations of Eqs. (A5.7.1) to (A5.7.3) are performed by applying the formula [2]

$$\int L_1^a L_2^b \, d\Omega_e = \frac{a!\,b!}{(a+b+1)!}\delta_x \qquad \text{(A5.7.5)}$$

The basis functions for the one-dimensional linear element are

$$N_1(x) = L_1 \quad \text{and} \quad N_2(x) = L_2 \qquad \text{(A5.7.6)}$$

Their derivates are given by

$$\nabla N_1(x) = \frac{dL_1}{dx} = \frac{d}{dx}\left(\frac{x_2^e - x}{\delta_x}\right) = \frac{0-1}{\delta_x} = -\frac{1}{\delta_x} \qquad \text{(A5.7.7)}$$

$$\nabla N_2(x) = \frac{dL_2}{dx} = \frac{d}{dx}\left(\frac{x - x_1^e}{\delta_x}\right) = \frac{1-0}{\delta_x} = -\frac{1}{\delta_x} \qquad \text{(A5.7.8)}$$

Denoting the conductive terms by the superscript e, the elements of the matrix $\mathbf{m}_e$ are obtained from Eq. (A5.7.5):

$$\begin{aligned}(m_e^d)_{11} &= \rho c_p A \int_{\Omega_e} N_1 N_2 \, d\Omega_e = \rho c_p A \int_{\Omega_e} L_1 L_2 \, d\Omega_e = \rho c_p A \int_{\Omega_e} L_1 L_1 \, d\Omega_e \\ &= \rho c_p A \int_{\Omega_e} L_1^{\,2} \, d\Omega_e = \rho c_p A \int_{\Omega_e} L_1^{\,2} L_2^{\,0} \, d\Omega_e = \rho c_p A \times \frac{2!\,0!}{(2+0+1)!}\delta_x \\ &= \rho c_p A \times \frac{2!}{3!} = \frac{1}{3}\rho c_p A \delta_x \end{aligned} \qquad \text{(A5.7.9)}$$

$$\left(m_e^d\right)_{12}=\rho c_\mathrm{p} A\int_{\Omega_e} N_1 N_2\, d\Omega_e=\rho c_\mathrm{p} A\int_{\Omega_e} L_1 L_2\, d\Omega_e=\rho c_\mathrm{p} A\int_{\Omega_e} L_1{}^1 L_2{}^1 d\Omega_e$$

$$=\rho c_\mathrm{p} A\frac{1!\,1!}{(1+1+1)!}\delta_x=\frac{1}{3!}\rho c_\mathrm{p} A\delta_x=\frac{1}{3\times 2\times 1}\rho c_\mathrm{p} A\delta_x=\frac{1}{6}\rho c_\mathrm{p} A\delta_x \tag{A5.7.10}$$

$$\left(m_e^d\right)_{21}=\rho c_\mathrm{p} A\int_{\Omega_e} N_2 N_1\, d\Omega_e=\rho c_\mathrm{p} A\int_{\Omega_e} L_2 L_1\, d\Omega_e=\rho c_\mathrm{p} A\int_{\Omega_e} L_2{}^1 L_1{}^1 d\Omega_e$$

$$=\rho c_\mathrm{p} A\frac{1!\,1!}{(1+1+1)!}\delta_x=\frac{1}{3!}\rho c_\mathrm{p} A\delta_x=\frac{1}{3\times 2\times 1}\rho c_\mathrm{p} A\delta_x=\frac{1}{6}\rho c_\mathrm{p} A\delta_x \tag{A5.7.11}$$

$$\left(m_e^d\right)_{22}=\rho c_\mathrm{p} A\int_{\Omega_e} N_1 N_2\, d\Omega_e=\rho c_\mathrm{p} A\int_{\Omega_e} L_1 L_2\, d\Omega_e=\rho c_\mathrm{p} A\int_{\Omega_e} L_2 L_2 d\Omega_e$$

$$=\rho c_\mathrm{p} A\int_{\Omega_e} L_2{}^2 d\Omega_e=\rho c_\mathrm{p} A\int_{\Omega_e} L_2{}^2 L_1{}^0 d\Omega_e$$

$$=\rho c_\mathrm{p} A\frac{2!\,0!}{(2+0+1)!}\delta_x=\rho c_\mathrm{p} A\times\frac{2!}{3!}=\frac{1}{3}\rho c_\mathrm{p} A\delta_x \tag{A5.7.12}$$

Combining together the elements of the matrix $\mathbf{m}_e^d$ given in Eqs. (A5.7.9)–(A5.7.12), we have

$$m_e^d=\begin{bmatrix}\left(m_e^d\right)_{11} & \left(m_e^d\right)_{12}\\ \left(m_e^d\right)_{21} & \left(m_e^d\right)_{22}\end{bmatrix}=\begin{bmatrix}\frac{1}{3}\rho c_\mathrm{p} A\delta_x & \frac{1}{6}\rho c_\mathrm{p} A\delta_x\\ \frac{1}{6}\rho c_\mathrm{p} A\delta_x & \frac{1}{3}\rho c_\mathrm{p} A\delta_x\end{bmatrix}=\rho c_\mathrm{p} A\delta_x\begin{bmatrix}\frac{1}{3} & \frac{1}{6}\\ \frac{1}{6} & \frac{1}{3}\end{bmatrix} \tag{A5.7.13}$$

Similarly, the elements of the matrix $\mathbf{k}_e$ are calculated as follows:

$$\left(k_e^d\right)_{11}=kA\int_{\Omega_e}\nabla N_1\cdot\nabla N_1\, d\Omega_e=kA\int_{\Omega_e}\left(-\frac{1}{\delta_x}\right)\cdot\left(-\frac{1}{\delta_x}\right)d\Omega_e$$

$$=\frac{kA}{(\delta_x)^2}\int_{\Omega_e} L_1{}^0 L_2{}^0 d\Omega_e=\frac{kA}{(\delta_x)^2}\frac{0!\,0!}{(0+0+1)!}\delta_x=\frac{kA}{\delta_x} \tag{A5.7.14}$$

$$(k_e^d)_{12} = kA\int_{\Omega_e} \nabla N_1 \cdot \nabla N_2 \, d\Omega_e = kA\int_{\Omega_e}\left(-\frac{1}{\delta_x}\right)\cdot\left(\frac{1}{\delta_x}\right) d\Omega_e$$

$$= \frac{kA}{(\delta_x)^2}\int_{\Omega_e} L_1{}^0 L_2{}^0 \, d\Omega_e = -\frac{kA}{(\delta_x)^2}\frac{0!\,0!}{(0+0+1)!}\delta_x = -\frac{kA}{\delta_x} \qquad \text{(A5.7.15)}$$

$$(k_e^d)_{21} = kA\int_{\Omega_e} \nabla N_2 \cdot \nabla N_1 \, d\Omega_e = kA\int_{\Omega_e}\left(\frac{1}{\delta_x}\right)\cdot\left(\frac{-1}{\delta_x}\right) d\Omega_e$$

$$= \frac{kA}{(\delta_x)^2}\int_{\Omega_e} L_1{}^0 L_2{}^0 \, d\Omega_e = \frac{kA}{(\delta_x)^2}\frac{0!\,0!}{(0+0+1)!}\delta_x = -\frac{kA}{\delta_x} \qquad \text{(A5.7.16)}$$

$$(k_e^d)_{22} = kA\int_{\Omega_e} \nabla N_2 \cdot \nabla N_2 \, d\Omega_e = kA\int_{\Omega_e}\left(\frac{1}{\delta_x}\right)\cdot\left(\frac{1}{\delta_x}\right) d\Omega_e$$

$$= \frac{kA}{(\delta_x)^2}\int_{\Omega_e} L_1{}^0 L_2{}^0 \, d\Omega_e = \frac{kA}{(\delta_x)^2}\frac{0!\,0!}{(0+0+1)!}\delta_x = \frac{kA}{\delta_x} \qquad \text{(A5.7.17)}$$

Combining together the elements of the matrix $\mathbf{k}_e^d$ given in Eqs. (A5.7.14)–(A5.7.17), we get

$$\mathbf{k}_e^d = \begin{bmatrix} (k_e^d)_{11} & (k_e^d)_{12} \\ (k_e^d)_{21} & (k_e^d)_{22} \end{bmatrix} = \begin{bmatrix} \dfrac{kA}{\delta_x} & -\dfrac{kA}{\delta_x} \\ -\dfrac{kA}{\delta_x} & \dfrac{kA}{\delta_x} \end{bmatrix} = \frac{kA}{\delta_x}\begin{bmatrix} 1 & -1 \\ -1 & 1 \end{bmatrix} \qquad \text{(A5.7.18)}$$

Realization of the matrix $\mathbf{m}_e^d$ is implemented by means of three capacitors, obtained as [1]

$$C_1^d = C' = (m_e^d)_{11} + (m_e^d)_{12} = \tfrac{1}{3}\rho c_{\mathrm{p}} A\delta_x + \tfrac{1}{6}\rho c_{\mathrm{p}} A\delta_x = \tfrac{1}{2}\rho c_{\mathrm{p}} A\delta_x \qquad \text{(A5.7.19)}$$

$$C_2^d = C = -(m_e^d)_{12} = -\tfrac{1}{6}\rho c_{\mathrm{p}} A\delta_x \qquad \text{(A5.7.20)}$$

$$C_3^d = C' = (m_e^d)_{21} + (m_e^d)_{22} = \tfrac{1}{6}\rho c_{\mathrm{p}} A\delta_x + \tfrac{1}{3}\rho c_{\mathrm{p}} A\delta_x = \tfrac{1}{2}\rho c_{\mathrm{p}} A\delta_x \qquad \text{(A5.7.21)}$$

Likewise, realization of the matrix $\mathbf{k}_e^d$ is achieved by a resistor [1]

$$R^d = R = -\frac{1}{(k_e^d)_{12}} = \frac{\delta_x}{kA} \qquad \text{(A5.7.22)}$$

Furthermore, because

$$\frac{1}{\left(k_e^d\right)_{11} + \left(k_e^d\right)_{12}} = \frac{1}{\dfrac{kA}{\delta_x} + \left(-\dfrac{kA}{\delta_x}\right)} = \frac{1}{0} = \infty \tag{A5.7.23}$$

$$\frac{1}{\left(k_e^d\right)_{21} + \left(k_e^d\right)_{22}} = \frac{1}{\left(-\dfrac{kA}{\delta_x}\right) + \dfrac{kA}{\delta_x}} = \frac{1}{0} = \infty \tag{A5.7.24}$$

no resistance is connected between the two element nodes and ground. Thus the equivalent conductive circuit network for the one-dimensional linear element is constructed.

REFERENCES

1. J. T. Hsu and L. Vu-Quoc, A Rational Formulation of Thermal Circuit Models for Electrothermal Simulation—Part I: Finite Element Method, *IEEE Trans. Circuits & Systems—*I: *Fundamental Theory and Applications*, Vol. 43, No. 9, September 1996, pp. 721–732.
2. O. C. Zienkiewicz and R. L. Taylor, *The Finite Element Method*, Vol. 1, 4th Edition, McGraw Hill, New York, 1989.

6

LATCHUP OF PARASITIC THYRISTOR IN IGBT

6.1 INTRODUCTION

A "latch" is a fastening for a gate or door. When latching takes place, the IGBT is tied to a state of continuous current conduction. There is, then, no influence of applied gate voltage on output collector current. The family of output collector current characteristics of the IGBT at different gate voltages reduces to a single curve independent of gate voltage. The phenomenon is observed as a decrease of forward voltage drop of the IGBT. Thus, it may be defined as a high current state for the IGBT, accompanied by a collapsing or low-voltage condition. Cessation of latching is brought about only by reversal of polarity of the collector voltage or by switching off this voltage. In DC or AC applications, the heat generated in the device may be enormous, causing its burnout.

Two distinct modes of latching in IGBT have been identified: static and dynamic. *Static latching* occurs when the IGBT is in its forward conduction state. It happens when the current density exceeds a limiting critical value. So in this mode, the collector voltage is low but the collector current is high. Nondestructive tests on IGBTs have revealed that the latching process is not confined to local regions in the IGBT, but spreads over most of its active area. *Dynamic latching* occurs during switching conditions when both the collector current and collector voltage are high. But the current density required for dynamic latching is lower than that for the static mode. In AC

The Insulated Gate Bipolar Transistor (IGBT) Theory and Design, By Vinod Kumar Khanna.
ISBN 0-471-23845-7 © 2003 Institute of Electrical and Electronics Engineers

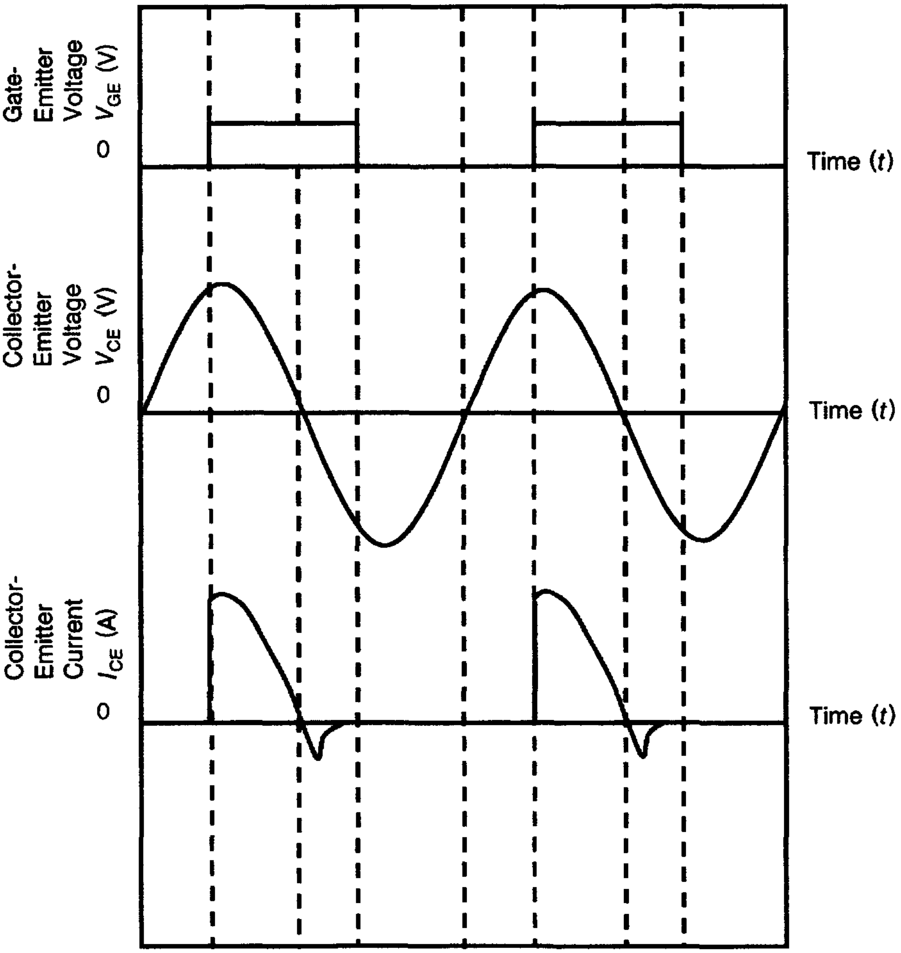


Figure 6.1 Gate and collector–emitter voltage/collector–emitter current switching waveforms for IGBT undergoing reverse recovery by the action of collector–emitter voltage (i.e., without forced gate–emitter voltage turn-off).

circuits (Fig. 6.1), applying a gate–emitter voltage during the positive half-cycle of the collector–emitter voltage waveform turns on the IGBT. Now, to switch off the IGBT, it is not necessary to remove the gate–emitter voltage, since during the negative half-cycle of collector–emitter voltage the direction of this voltage reverses and the IGBT undergoes reverse recovery. Turning off the IGBT in this manner, without withdrawal of the gate–emitter voltage, the dynamic latching current density equals the static latching current density. If now the gate–emitter voltage is removed during IGBT turn-off, the dynamic current density of latching will be less than that for the static case. The decrease in latching current density, arising from gate-controlled turn-off during dynamic operation, is in sharp distinction from the latching encountered during steady-state operation of the device.

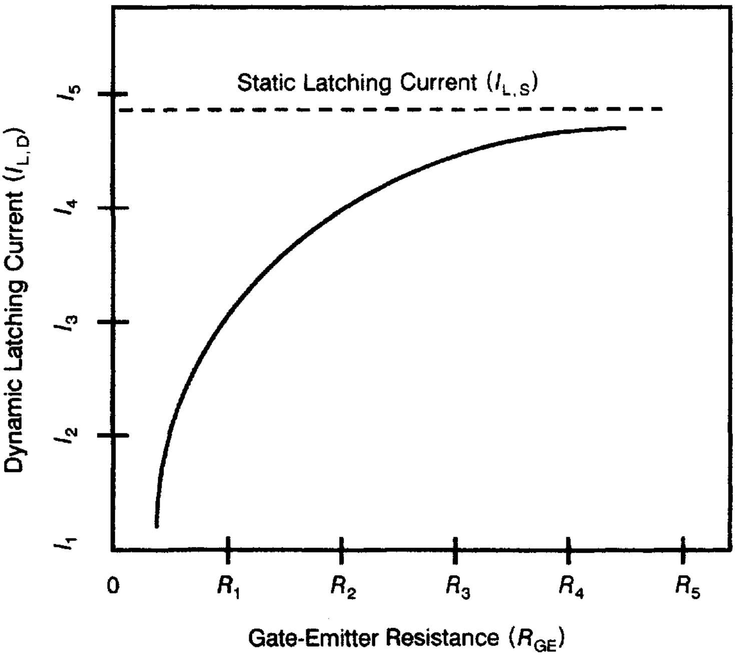


Figure 6.2 Increase in dynamic latching current of IGBT with gate–emitter series resistance ($R_5 > R_4 > R_3, \ldots,$). For higher values of R_{GE}, the dynamic latching current approaches the static limit.

During IGBT turn-off, the rapidly disappearing internal MOSFET action causes a heavy hole flux across the base resistance R_b in the reverse direction, creating a voltage drop. When this voltage drop exceeds 0.7 V, a high injection of electrons occurs from the N^+ emitter into the P base. Thus, in the dynamic mode, latching takes place with lower collector current values due to the hole overcurrent. Common usage of IGBTs in circuits demanding forced-gate turn-off calls for ensuring that the dynamic latching current be kept far above the maximum operating current. Because static latching current is less than the dynamic latching current, imposition of this design constraint will serve to avoid static latching. It may therefore be viewed as a *worst-case design criterion*, which, if fulfilled, will guarantee nonlatching operation, under all operating conditions of the device.

It may be mentioned here that if the turn-off process can be slowed down, a major fraction of the inrushing holes will recombine in the drift zone Then the chances of latching will be relatively small. An easy circuit solution of this problem is to use a larger value of series gate–emitter resistance R_{GE} to reduce the speed of turn-off (Fig. 6.2). Also, a smaller gate–emitter voltage (15 V) keeps the collector current within reasonable limits and lessens the likelihood of latching. Other latching prevention techniques call for suitable modifications of doping profiles, along with geometrical and dimensional adjustments, during device design [1]. These will be discussed in Section 6.4.

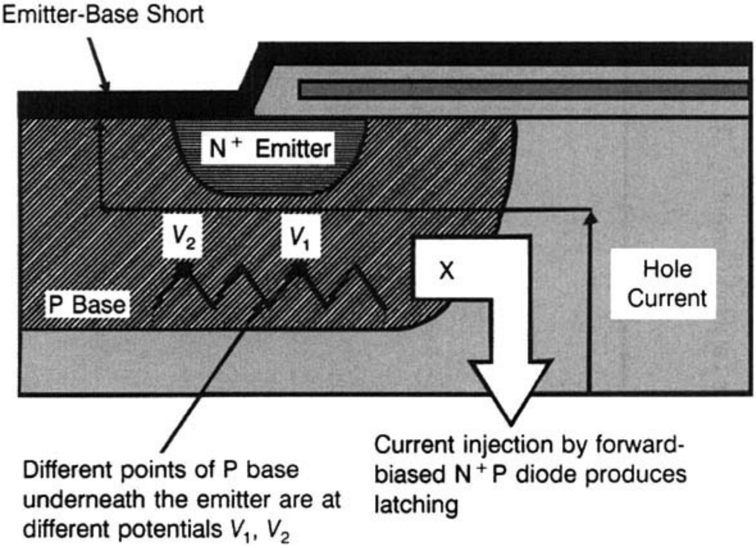

Emitter-Base Short
N⁺ Emitter
V₂
V₁
P Base
X
Hole Current
Different points of P base underneath the emitter are at different potentials V₁, V₂
Current injection by forward-biased N⁺P diode produces latching

Figure 6.3 Region of IGBT responsible for initiation of latching.

6.2 STATIC LATCHING

Under static latching, the upper NPN transistor in the IGBT is no longer inactive. The hole current I_p , flowing laterally into the P base, traverses a path under the N^+ emitter and is collected by the emitter metal at the emitter/P-base short (Fig. 6.3). This laterally flowing hole current forward biases the N^+-emitter/P-base junction at a point X. Although this hole collection takes place along the full length of the P base, to calculate the forward drop across the N^+-P junction, an effective hole current moving under the entire emitter length L_E may be used. If R_p is the resistance of P base under the N^+ emitter, we obtain

$$V_X = R_P I_p \tag{6.1}$$

But we know [see Eq. (5.20)] that the hole current (I_p) is related to the electron current (I_n) flowing through the MOSFET channel as

$$I_p = \frac{\alpha_{PNP} I_n}{(1 - \alpha_{PNP})} \tag{6.2}$$

where α_{PNP} is the current gain of PNP transistor. Also, emitter current is given by

$$I_E = I_p + I_n = \frac{I_n}{(1 - \alpha_{PNP})} \tag{6.3}$$

Applying Eqs. (6.2) and (6.3) to Eq. (6.1), we get

$$V_X = \alpha_{PNP} R_P I_{CE} \tag{6.4}$$

Latching occurs when the sum-total current gain ($\alpha_{NPN} + \alpha_{PNP}$) of the constituent transistors attains a value of unity [Eq. (4.70)], that is,

$$\alpha_{NPN} + \alpha_{PNP} = 1 \tag{6.5}$$

For increasing the current gain of the NPN transistor, to the value at which this condition is validated, the forward bias on the N^+-P junction must equal the built-in potential ($V_{bi} = 0.7$ V) of this junction. Hence, the steady-state latching current is expressed a

$$I_{L,SS} = \frac{V_{bi}}{\alpha_{PNP} R_P} \tag{6.6}$$

But

$$R_P \text{ is proportional to } (\propto)\ \rho_{sp} L_E \tag{6.7}$$

where $\rho_{s,p}$ is the sheet resistance of the P base. Therefore,

$$I_{L,ss} \propto \frac{V_{bi}}{\alpha_{PNP}\, \rho_{sp} L_E} \tag{6.8}$$

or, simply stated, the necessary conditions for steady-state latching of the IGBT are $\alpha_{NPN} + \alpha_{PNP} \geq 1$, wherein the IGBT could perform latchup like a thyristor. Equation (6.8) suggests the solutions to the latching problem. Latching current can be raised by (i) decreasing the current gain of the PNP transistor (α_{PNP}), (ii) decreasing the sheet resistance ($\rho_{s,p}$) of the P base, (iii) using a shorter total emitter length L_E, and (iv) a combination of the above three methods.

Example 6.1 In an IGBT, the current gain of PNP transistor was 0.5. P-base concentration was 9×10^{16} cm^{-3}, and diffusion depth was 4 μm. Emitter diffusion was carried out through a 5-μm window up to a depth of 1 μm. The latching current density of the device was measured to be 900 A/cm^2. If now the P-base concentration is altered to 1.5×10^{17} cm^{-3} and the emitter diffusion window is taken as 2 μm, what will be the effect on the latching current density?

The IGBT half-cell for this example is shown in Fig. E6.1.1. In the first case, Common-base current gain of the PNP transistor, α_{PNP}, is 0.5, sheet resistance of P base, $\rho_{s,p}$, is expressed as Resistivity/junction depth = {1/(Carrier concentration × Electronic charge × Mobility)}/junction depth = $\{1/(9 \times 10^{16} \times 1.6 \times 10^{-19} \times 480)\}/4 \times 10^{-4} = 361.69\ \Omega/\square$. Taking into account that the lateral diffusion is 80% of the vertical junction depth, emitter length $L_E = 5 + 2 \times 0.8 = 6.6$ μm. Static

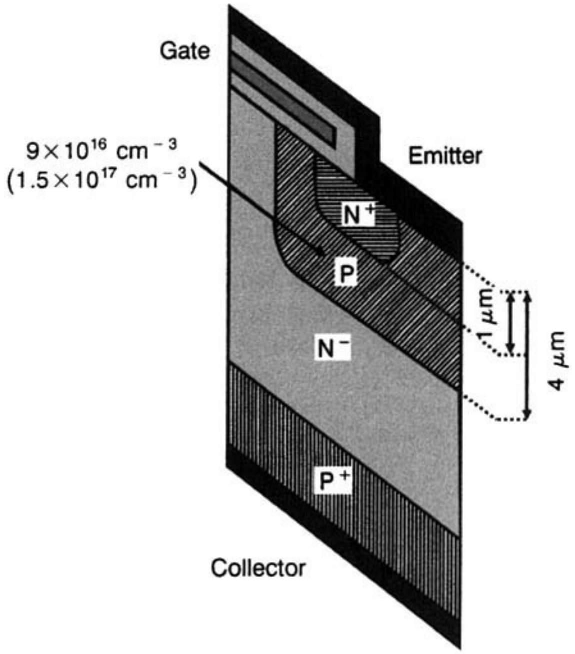


Figure E6.1.1 IGBT for Example 6.1.

latching current density, $J_{L,ss}$ is 900 A/cm^2. From Eq. (6.8), $900 = k/0.5 \times 361.69 \times 6.6 \times 10^{-4}$, where k is a proportionality constant that includes V_{bi}. This equation gives $k = 107.42$. In the second case, $\rho_{s,p} = \{1/(1.5 \times 10^{17} \times 1.6 \times 10^{-19} \times 480)\}/4 \times 10^{-4} = 217\ \Omega/\square$. $L_E = 2 + 2 \times 0.8 = 3.6\ \mu$m $= 3.6 \times 10^{-4}$ cm. Therefore, $J_{L,ss} = 107.42/0.5 \times 217 \times 3.6 \times 10^{-4}$ equals 2750.12 A/cm^2.

6.3 DYNAMIC LATCHING

6.3.1 Latching of Symmetrical IGBTs with Resistive Load

Recall the discussions on symmetrical and asymmetrical IGBTs in Section 2.1.2. In a symmetrical IGBT, doping density distribution in N$^-$ base is homogeneous. It has the layer structure: P$^+$ substrate/N$^-$ base/P base/N$^+$ emitter. The current and voltage waveforms for resistive load turn-off in a DC circuit (Fig. 6.4a) are drawn in Fig. 6.4b for a small gate–emitter resistance R_{GE}. At the instant $t = t_1$, the collector–emitter voltage, V_{CE}, increases sharply from the forward drop V_F of IGBT to a value V_1 that is determined by the magnitude of collector–emitter current fall ΔI_{CE}. Since

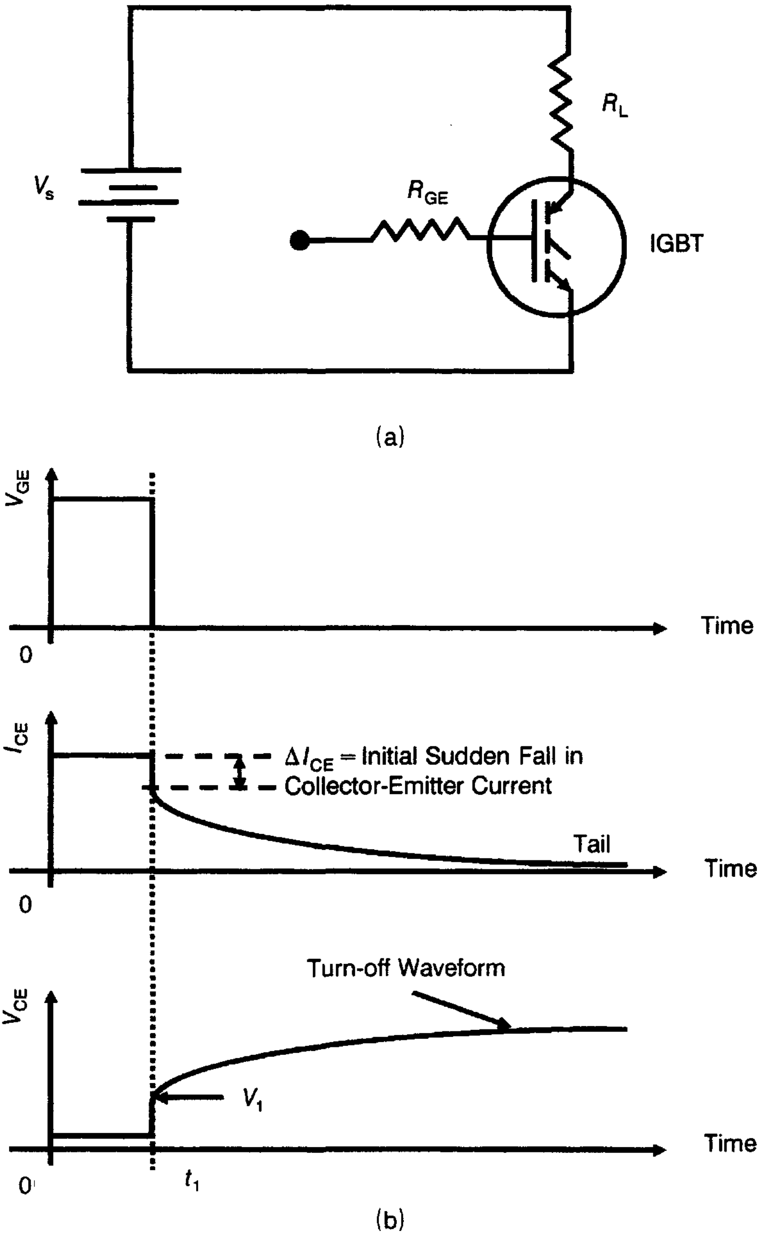

R_L
V_s
R_{GE}
IGBT
(a)
V_{GE}
Time
0
I_{CE}
ΔI_{CE} = Initial Sudden Fall in Collector-Emitter Current
Tail
Time
0
Turn-off Waveform
V_{CE}
V_1
Time
0
t_1
(b)

Figure 6.4 Forced gate turn-off of IGBT with resistive load: (a) Circuit and (b) waveform.

$V_F \ll$ supply voltage V_s, we may write

$$V_1 = \Delta I_{CE} R_L \tag{6.9}$$

As this voltage is applied to the depletion region across the junction between P base and N^- base, the depletion region spreads mainly into the low-doped N^- base. This shortens the undepleted base width of the PNP transistor in the IGBT and thereby increases its current gain α_{PNP} at time instant t_1 corresponding to turn-off. So, when turn-off occurs at $t = t_1$, the current gain is [Appendix 4.8, Eq. (A4.8.17)]

$$\alpha_{PNP,DR} = \frac{1}{\cosh(W_{DR}/L_a)} \tag{6.10}$$

where D in the subscript indicates dynamic turn-off and R stands for resistive load. The undepleted base width in this transistor is

$$W_{DR} = 2d - \sqrt{\frac{2\varepsilon_0 \varepsilon_s V_1}{qN_D}} \tag{6.11}$$

where $2d$ is the N^--base thickness. This is because

Undepleted base width W_{DR} = Total base thickness − Depletion region width

Hence, applying Eq. (4.13), we obtain

$$W_{DR} = 2d - \sqrt{\frac{2\varepsilon_0 \varepsilon_s(\phi_0 - v_a)(N_A + N_D)}{qN_A N_D}} \cong 2d - \sqrt{\frac{2\varepsilon_0 \varepsilon_s V_1 (N_A + N_D)}{qN_A N_D}}$$

$$= 2d - \sqrt{\frac{2\varepsilon_0 \varepsilon_s V_1 \left(\dfrac{1}{N_D} + \dfrac{1}{N_A}\right)}{q}} \cong 2d - \sqrt{\frac{2\varepsilon_0 \varepsilon_s V_1 \times \dfrac{1}{N_D}}{q}}$$

where $\phi_0 - v_a$ has been replaced by V_1, and both numerator and denominator have been divided by $N_A N_D$. Also, $(1/N_D) \gg (1/N_A)$ for P^+N diode as $N_D \ll N_A$.

However, prior to turn-off at $t < t_1$, the current gain of the PNP transistor was

$$\alpha_{PNP} = \frac{1}{\cosh(2d/L_a)} = 2\exp(-2d/L_a) \tag{6.12}$$

because the IGBT is in the forward-biased conducting state. Hence the depletion layer width at the small forward bias is insignificant. Justification for Eq. (6.12) is as follows:

$$\frac{1}{\cosh(2d/L_{\mathrm{s}})} = \frac{1}{\dfrac{\exp(2d/L_{\mathrm{a}}) + \exp(-2d/L_{\mathrm{a}})}{2}}$$

$$= \frac{2}{\exp(2d/L_{\mathrm{a}}) + \exp(-2d/L_{\mathrm{a}})}$$

We take typical values of parameters such as $d = 100$ μm, $D_{\mathrm{n}} = 34.96$ cm^2/sec, $D_{\mathrm{p}} = 12.43$ cm^2/sec, and $D_{\mathrm{a}} = 2 \times 34.96 \times 12.43/(34.96 + 12.43) =$ 18.34 cm^2/sec, so that for carrier lifetime we obtain $\tau_{\mathrm{a}} = 5$ μsec, $L_{\mathrm{a}} = \sqrt{(D_{\mathrm{a}}\tau_{\mathrm{a}})} = \sqrt{(18.34 \times 5 \times 10^{-6})} = 0.0096$ cm. Then $\exp(2d/L_{\mathrm{a}}) = \exp(2 \times 100 \times 10^{-4}/0.0096) = 8.03$, and $\exp(-2d/L_{\mathrm{a}}) = 0.12$. But for $\tau_{\mathrm{a}} = 10$ μsec, $L_{\mathrm{a}} =$ 0.01354 cm. So, $\exp(2d/L_{\mathrm{a}}) = 4.38$ and $\exp(-2d/L_{\mathrm{a}}) = 0.228$. However, the ratio $\{\exp(2d/L_{\mathrm{a}})\}/\{\exp(-2d/L_{\mathrm{a}})\}$ decreases from 66.91 to 19.21 on increasing the lifetime from 5 to 10 μsec, but still $\exp(2d/L_{\mathrm{a}})$ remains much greater than $\exp(-2d/L_{\mathrm{a}})$. Hence,

$$\frac{1}{\cosh(2d/L_{\mathrm{a}})} \cong \frac{2}{\exp(2d/L_{\mathrm{a}})} = 2\exp(-2d/L_{\mathrm{a}})$$

Now Eq. (6.10) represents the increased current gain after turn-off ($t \geq t_1$) while Eq. (6.12) gives the lower current gain before turn-off ($t < t_1$). Despite the abrupt rise of current gain at $t \geq t_1$, the hole current flowing in the P base after turn-off remains the same as before turn-off. Latching occurs when

$$\alpha_{\mathrm{NPN,DR}} + \alpha_{\mathrm{PNP,DR}} \geq 1 \tag{6.13}$$

or,

$$\alpha_{\mathrm{NPN,DR}} = 1 - \alpha_{\mathrm{PNP,DR}} \tag{6.14}$$

Because $\alpha_{\mathrm{PNP,\,DR}} > \alpha_{\mathrm{PNP,\,SS}}$, it is evident that hole current I_{h} at which latching takes place, and therefore the steady-state collector current I_{CE} before turn-off, is less than that for the case of static latching. This is why dynamic latching occurs at a lower value of collector-emitter current. Postulating that α_{NPN} and I_{h} are linearly related, we can immediately write the ratio of dynamic and static latching currents as

$$I_{\mathrm{L,DR}}/I_{\mathrm{L,SS}} = I_{\mathrm{p,DR}}/I_{\mathrm{p,SS}} = \alpha_{\mathrm{NPN,DR}}/\alpha_{\mathrm{NPN,SS}} = (1 - \alpha_{\mathrm{PNP,DR}})(1 - \alpha_{\mathrm{PNP,SS}})$$

$$= \frac{\exp(2d/L_{\mathrm{a}}) - 2\exp\left(\sqrt{2\varepsilon_0\varepsilon_{\mathrm{s}}V_1/qN_{\mathrm{D}}L_{\mathrm{a}}^{\,2}}\right)}{\exp(2d/L_{\mathrm{a}}) - 2} \tag{6.15}$$

$$= \frac{\exp(2d/L_{\mathrm{a}}) - 2\exp\left(\sqrt{2\varepsilon_0\varepsilon_{\mathrm{s}}\Delta I_{\mathrm{CE}}R_{\mathrm{L}}/qN_{\mathrm{D}}L_{\mathrm{a}}^{\,2}}\right)}{\exp(2d/L_{\mathrm{a}}) - 2}$$

upon substituting for V_1 from Eq. (6.9). Equation (6.15) is derived in Appendix 6.1.

Noting that ΔI_{CE} depends on α_{PNP} during steady-state forward conduction, we have

$$\Delta I_{\mathrm{CE}} = I_{\mathrm{n}} = (1 - \alpha_{\mathrm{PNP,SS}}) I_{\mathrm{CE}} \tag{6.16}$$

This equation is obvious from Eq. (5.13) because $I_{\mathrm{CE}} \cong I_{\mathrm{E}}$. Also, neglecting the forward voltage drop across the IGBT, we may write

$$V_{\mathrm{s}} = I_{\mathrm{CE}} R_L \tag{6.17}$$

Including these modifications, the ratio of latching currents becomes

$$\frac{I_{\mathrm{L,DR}}}{I_{\mathrm{L,SS}}} = \frac{\exp(2d/L_{\mathrm{a}}) - 2\exp\left\{\sqrt{2\varepsilon_0 \varepsilon_{\mathrm{s}}(1 - \alpha_{\mathrm{PNP,SS}}) V_{\mathrm{s}}/qN_{\mathrm{D}} L_{\mathrm{a}}^{\,2}}\right\}}{\exp(2d/L_{\mathrm{a}}) - 2} \tag{6.18}$$

As the DC supply voltage V_{s} increases, the latching current ratio increases. So, for the higher supply voltages, the IGBT will be more prone to dynamic latching. Reduction of the PNP transistor current gain, as happens during electron irradiation to increase the switching speed, has the effect of increasing the susceptibility to dynamic current latching. This is contrary to the thinking that any decrease of current gains of the bipolar transistors will be beneficial to latching control. In reality, dynamic latching current will be lower after subjecting the IGBT to lifetime killing. This is borne out by experimental investigations.

Example 6.2 Calculate the ratio of dynamic latching current to static latching current for an NPT-IGBT in a resistive load DC circuit using a supply voltage of 400 V, given that the N$^-$-base dopant density is 1×10^{14} cm^{-3}, width is 200 μm, ambipolar lifetime in N$^-$ base is 10 μsec, and electron and hole diffusion coefficients are 35 and 12.4 cm^2/sec. What will happen if the DC supply voltage is 600 V? What will be the effect if the ambipolar lifetime is killed down to 1 μsec but the power supply voltage is kept at 400 V?

Figure E6.2.1 shows the IGBT half-cell. Ambipolar diffusion coefficient D_{a} is

$$D_{\mathrm{a}} = 2D_{\mathrm{n}}D_{\mathrm{p}}/(D_{\mathrm{n}} + D_{\mathrm{p}}) \tag{E6.2.1}$$

where D_{n} and D_{p} are electron and hole diffusion coefficients. This equation gives $D_{\mathrm{a}} = 18.3$ cm^2/sec. Therefore ambipolar diffusion length L_{a} is expressed as

$$L_{\mathrm{a}} = \sqrt{(D_{\mathrm{a}}\tau_{\mathrm{a}})} \tag{E6.2.2}$$

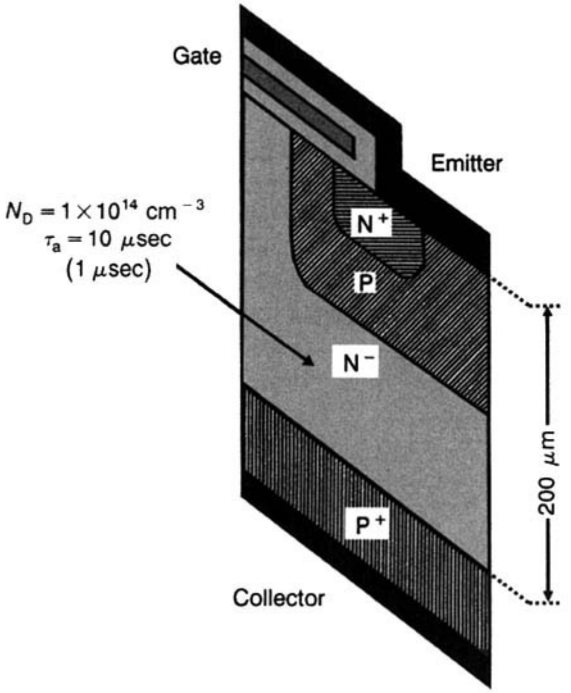


Figure E6.2.1 IGBT for Example 6.2.

where τ_a is the carrier lifetime. Thus $L_a = \sqrt{(18.3 \times 10 \times 10^{-6})} = 0.01353$ cm. The PNP transistor current gain α_{PNP} is given by

$$\alpha_{PNP} = 1/\cosh(2d/L_a) \qquad \text{(E6.2.3)}$$

where $2d$ is the N-base width. Here it is assumed that the emitter injection efficiency is unity. For the IGBT in question, we obtain $\alpha_{PNP} = 1/\cosh(200 \times 10^{-4}/0.01353) = 0.4335$. Also, it is given that $V_s = 400$ V. Substitution of the above values in Eq. (6.18) yields $I_{L,DR}/I_{L,SS} = [\exp(200 \times 10^{-4}/0.01353) - 2\exp\sqrt{\{2 \times 8.854 \times 10^{-14} \times 11.9 \times (1 - 0.4335) \times 400/(1.6 \times 10^{-19} \times 1 \times 10^{14} \times 0.01353^2)\}}]/\{\exp(200 \times 10^{-4}/0.01353) - 2\} = (4.385 - 2.9949)/(4.385 - 2) = 1.3901/2.385 = 0.583$.

When $V_s = 600$ V, we have $I_{L,DR}/I_{L,SS} = (4.385 - 3.2794)/2.385 = 0.4636$. The latching ratio worsens and the dynamic latching occurs at a lower current density with a power supply of 600 V than while using 400 V.

If the carrier lifetime is reduced to 1 μsec, $L_a = 4.28 \times 10^{-3}$ cm, $\alpha_{PNP} = 0.0187$, $I_{L,DR}/I_{L,SS} = (107.007 - 10.73)/(107.007 - 2) = 0.9168$. Because of the fall in the gain of PNP transistor, lifetime killing makes the decline of dynamic latching current density less severe.

6.3.2 Latching of Asymmetric IGBTs with Resistive Load

An asymmetrical IGBT has the following layer structure: P^+ substrate/N^+ buffer layer/N^- base/P base/N^+ emitter. Here the N^--base region comprises two portions: a higher resistivity N^- layer of thickness d_{N-} and a lower resistivity N-buffer layer of thickness d_N. Referring to Fig. 6.4, at any value of voltage V_1, the N^- layer is completely depleted and some portion of N-buffer layer is also depleted. Assuming that the depletion layer width in steady-state forward conduction is negligible, we can derive the ratio of dynamic to static latching currents following the procedure for a symmetrical IGBT. This analysis yields the equation

$$\frac{I_{L,DR}}{I_{L,SS}} = \frac{1 - \dfrac{1}{\cosh(d_N/L_a)}}{1 - \dfrac{1}{\cosh\{(d_{N^-} + d_N)/L_a\}}} \tag{6.19}$$

For achieving a high forward breakdown voltage, d_N and d_{N^-} are much larger than the diffusion length. Then (Appendix 6.2)

$$\frac{I_{L,DR}}{I_{L,SS}} = \frac{\exp(d_N/L_a) - 2}{\exp(d_N/L_a) - 2\exp(-d_{N^-}/L_a)} \tag{6.20}$$

From this equation it is readily apparent that the forced gate turn-off, with a resistive load, using an asymmetric IGBT is more prone to latching than the symmetric structure. This inference is drawn from the observation that even for V_1 (collector voltage at instant t_1) = V_s (supply voltage), the N^- layer thickness d_{N^-}, of asymmetrical IGBT is smaller than undepleted base width of symmetrical IGBT, producing a larger decrease of dynamic latching current for the asymmetric structure. Hence, the asymmetric IGBT is more likely to latch.

Example 6.3 A PT-IGBT comprises an N^--base region of thickness 40 μm and an N-buffer layer of thickness 10 μm. If the ambipolar diffusion length is 135.3 μm (same as in Example 6.2), estimate the dynamic to static latching current ratio for this structure. Compare this result with Example 6.2 and comment. What happens at a diffusion length = 35 μm.

The PT-IGBT half-cell is shown in Fig. E6.3.1. Here $d_{N^-} = 40 \times 10^{-4}$ cm and $d_N = 10 \times 10^{-4}$ cm, $L_a = 135.3 \times 10^{-4}$ cm. Applying Eq. (6.19), we get $I_{L,DR}/I_{L,SS} = [1 - 1/\{\cosh(10 \times 10^{-4}/135.3 \times 10^{-4})\}]/[1 - 1/\cosh\{(40 \times 10^{-4} + 10 \times 10^{-4})/135.3 \times 10^{-4}\}] = 2.725 \times 10^{-3}/0.065 = 0.0419$. Comparing this ratio with those

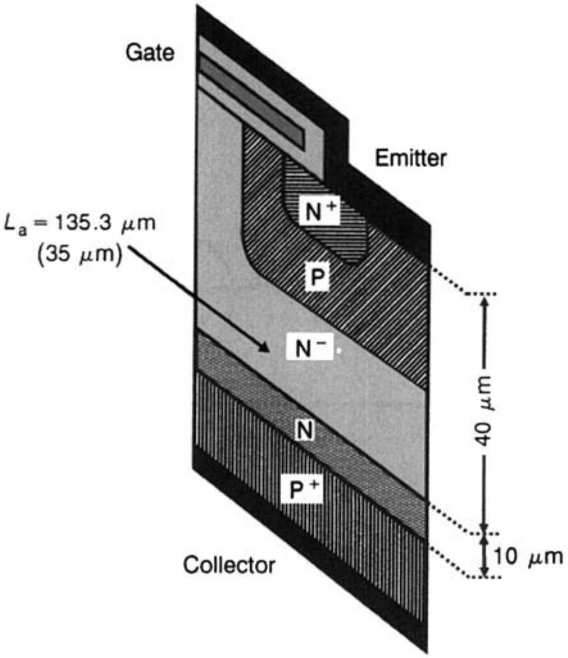


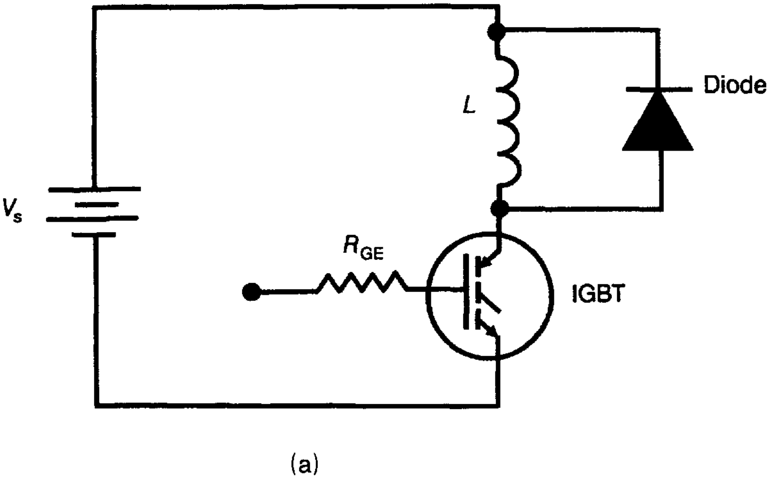
Figure E6.3.1 IGBT for Example 6.3.

obtained for the symmetric structures in Example 6.2, we find that the decrease in dynamic latching current density with reference to static mode is very severe for the asymmetric case. For $L_a = 35 \times 10^{-4}$ cm, we have $I_{L,DR}/I_{L,SS} = 0.03947/0.547 = 0.072$.

6.3.3 Latching of Symmetric IGBTs with Inductive Load

Figure 6.5a presents the inductive load turn-off circuit. The associated waveform is sketched in Fig. 6.5b. The collector–emitter voltage rises steeply to the supply voltage V_s at time instant $t = t_1$, slightly overshooting this voltage. Consequently, after turning off, the current gain of PNP transistor increases to a value corresponding to the voltage $V_1 = V_s$ so that

$$\alpha_{PNP,DI} = \frac{1}{\cosh(W_{DI}/L_a)} \tag{6.21}$$



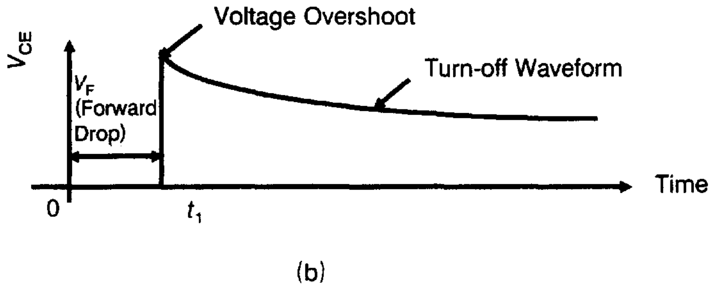


Figure 6.5 Inductive load turn-off of IGBT: (a) Circuit and (b) waveform.

where, as before, the symbol D is for dynamic and I is for inductive load.

$$W_{\mathrm{DI}} = 2d - \sqrt{\frac{\varepsilon_0 \varepsilon_s V_s}{q N_{\mathrm{D}}}} \tag{6.22}$$

is the undepleted base width for dynamic turn-off using inductive load. Equation (6.22) is similar to Eq. (6.11) with V_1 replaced by V_s. The analysis for this case is carried out in the same fashion as for the symmetrical IGBT, with resistive load giving the latching current ratio

$$\frac{I_{\mathrm{L,DI}}}{I_{\mathrm{L,SS}}} = \frac{1 - \alpha_{\mathrm{PNP,DI}}}{1 - \alpha_{\mathrm{PNP,SS}}} = \frac{\exp(2d/L_a) - 2\exp\left\{2\varepsilon_0 \varepsilon_s V_s / \left(q N_{\mathrm{D}} L_a^2\right)\right\}^{0.5}}{\exp(2d/L_a) - 2} \tag{6.23}$$

Comparing this equation with Eq. (6.18) for resistive load, we can easily see that the second term of the numerator of Eq. (6.18) contains the additional factor $(1 - \alpha_{\mathrm{PNP}})$, which is less than unity. Hence the second term of the numerator of Eq. (6.18) is smaller than the same for Eq. (6.23). So, the ratio of latching currents is larger for an inductive load, resulting in an increased propensity to latch. This implies that in forced gate inductive load turn-off, the IGBT will latch at a lower current than for resistive load turn-off. The influence of carrier lifetime on dynamic conditions is also different for the two types of loads. It must be pointed out that for resistive loads, lifetime killing reduces the dynamic latching current as compared to the static latching current. This means that for resistive loads the IGBT becomes more likely to latch under dynamic conditions after lifetime reduction. For inductive loads, the reverse remarks apply; that is, the chances of dynamic latching become less after lifetime degradation. This happens because the ratio of latching currents shows a weaker dependence on lifetime through the parameter L_a. But lifetime killing decreases the ratio of hole current I_h to collector–emitter current I_{CE}, thereby increasing the static latching current and hence the absolute value of dynamic latching current.

It must be clearly stated that in the latching current analysis of the above subsections, the impact of collector–emitter voltage on P-base resistance has been ignored. In reality, the depletion layer also stretches into the P base, although to a shorter distance. This raises the sheet resistance of P base and hence the resistance R_p. The effect becomes more severe at higher collector–emitter voltages. As a consequence, the dynamic latching will take place at lower currents than suggested by the above- simplified analysis. This effect can be accounted for, using numerical analysis.

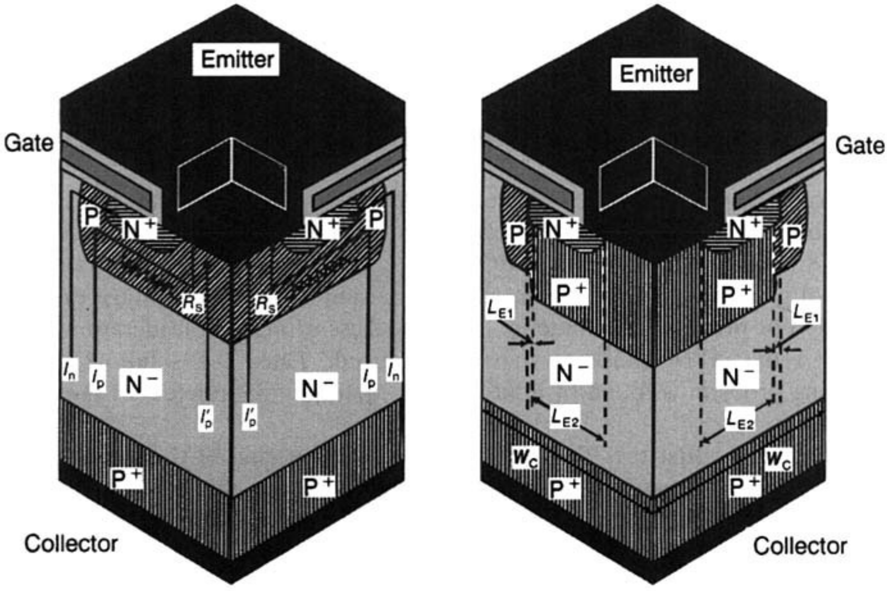
6.4 LATCHING PREVENTION MEASURES

The key to increase the latching current density lies in decreasing the current gain of either PNP transistor or NPN transistor or both transistors. As the PNP transistor amplifies the MOSFET channel current, reduction of its current gain has the undesirable effect of raising the forward voltage drop of IGBT. For proper IGBT functioning, regardless of other considerations, the gain of the NPN transistor must be reduced. Thus, the solution to the latching problem is found by adjusting the operating parameters of the NPN transistor.

Considering first the PNP transistor, the current gain of this transistor is lowered by decreasing its base transport factor α_T or the emitter injection efficiency γ. The base transport factor, α_T, declines during the electron irradiation of IGBT, which is essential to reduce turn-off time. During electron irradiation [2], the ambipolar lifetime τ_a falls and therefore the ambipolar diffusion length L_a in the N^--drift region decreases, resulting in a fall of base transport factor $\alpha_T = 1/\cosh(W/L_a)$. Reduction of emitter injection efficiency γ is accomplished by increasing the doping concentration of N$^-$-drift region. However, this has an adverse effect on the forward and reverse blocking capabilities of IGBT. Alternatively, the N-buffer layer (doping concentration $= 1 \times 10^{16}$ to 1×10^{17} cm^{-3}, and thickness = 10–15 μm) is incorporated in the IGBT structure [3]. The IGBT with the buffer layer provides comparable forward blocking capability to an IGBT, without the buffer layer, but the reverse blocking voltage is decreased. It also increases the ON-state voltage drop, but makes the IGBT faster switching.

Regarding the NPN transistor, which has received more attention for latching control, the commonly employed methods are summarized below:

1. Addition of Deep Heavily Doped (P$^+$) Region in the IGBT Structure. To decrease the resistance to hole current flow, the P-base doping concentration must be taken as 10^{19} cm^{-3}. On the other hand, for a threshold voltage of 2–3 V, the peak P-base doping concentration must be approximately 10^{17} cm^{-3}. In order to fulfill these conflicting requirements for the reduction of resistance to hole current without unduly increasing the threshold voltage, the P-base region under the emitter is made more conducting by an additional P$^+$ diffusion [4] which extends toward the MOSFET channel until the channel doping does not exceed 1×10^{17} cm^{-3}. Figure 6.6a shows an IGBT in which there is no P$^+$ diffused region under the emitter. The IGBT with P$^+$ diffused region below the emitter is shown in Figure 6.6b. With reference



(a) (b)

Figure 6.6 IGBT cell: (a) Without deep P$^+$ diffusion and (b) with deep P$^+$ diffusion.

to Fig. 6.6b, there are two segments under the N$^+$ emitter: (1) the segment of length L_{E1} containing the N$^+$ emitter over P-base region having a high sheet resistance of P-base $\rho_{SB,}$ and (2) the segment of length L_{E2}, containing the N$^+$ emitter over the P$^+$ diffusion having a low sheet resistance of P$^+$ diffusion ρ_{SP^+}. If Z is the width of the IGBT unit cell in the direction perpendicular to the cross section shown, the shunting resistance R_s is expressed as

$$R_s = (1/Z)(\rho_{SB} L_{E1} + \rho_{SP^+} L_{E2}) \tag{6.24}$$

This equation is obtained as follows: Component of resistance R_s arising from the P base = Resistivity of P base (ρ_B) × Distance L_{E1} traversed under the emitter/ Cross-sectional area for current flow = $(\rho_B \times L_{E1})/(X_{jP} \times Z)$, where X_{jP} is the junction depth of the P base. Then the component of R_s due to P base = $(\rho_B/X_{jP}) \times (L_{E1}/Z) = \rho_{sB} \times (L_{E1}/Z) = (1/Z)\rho_{sB} L_{E1}$, since sheet resistivity of P region ρ_{sB} = Bulk resistivity (ρ_B)/Junction depth X_{jP}. Similarly, the component of R_s due to P$^+$ deep diffusion in the P base is $(1/Z)\rho_{sP^+} L_{E2}$. Equation (6.24) represents the sum of the two components.

Latching takes place at a hole current component I_p, flowing under the N$^+$ emitter when its magnitude becomes

$$I_p = \frac{V_{bi}}{R_s} \tag{6.25}$$

where V_{bi} is the built-in potential of the N$^+$-emitter/P-base junction. This level of I_p determines the onset of latching because at this current the potential drop $I_p R_s$ equals the barrier or built-in potential (V_{bi}) of the N$^+$P diode, making it forward-biased (see Section 4.1.1). The forward-biased diode injects electrons into the N$^-$ base, thereby activating the NPN transistor of IGBT and increasing its current gain until the latching condition represented by Eq. (6.5) is satisfied. Neglecting the hole current component I_p flowing directly to the N$^+$-emitter/P-base short (region shorted by metal as marked in Fig. 6.3) without traversing the path underneath the emitter, we can write the relationship between the collector–emitter current I_{CE} and the hole component I_p as

$$I_{CE} = \frac{I_p}{\alpha_{PNP}} \tag{6.26}$$

where α_{PNP} is the current gain of the PNP transistor. This equation readily follows on considering the PNP transistor of the IGBT unit cell consisting of P$^+$ substrate, N$^-$ base, and P-base of IGBT; and bearing in mind that the emitter of this transistor is the collector of IGBT, we find the current gain of PNP transistor = α_{PNP} = Collector current of PNP transistor (or emitter current I_E of IGBT)/Emitter current of PNP transistor (or Collector current

I_{CE} of IGBT) $= I_E/I_{CE}$, or $I_{CE} = I_E/\alpha_{PNP} = I_p/\alpha_{PNP}$ since I_p has been taken as the total emitter current.

Combining the I_p and I_{CE} equations (6.25) and (6.26), using the R_s equation (6.24), and noting that the cross-sectional area for current flow is ZW_C (where W_C is the cell width), we obtain the latching current density:

$$J_{CEL} = \frac{I_{CEL}}{ZW_C} = \frac{\dfrac{I_p}{\alpha_{PNP}}}{ZW_C} = \frac{\dfrac{V_{bi}/R_s}{\alpha_{PNP}}}{ZW_C} = \frac{V_{bi}/(\alpha_{PNP}R_s)}{ZW_C}$$

$$= \frac{V_{bi}Z/\{\alpha_{PNP}(\rho_{SB}L_{E1} + \rho_{SP^+}L_{E2})\}}{ZW_C}$$

$$= \frac{V_{bi}}{W_C\alpha_{PNP}(\rho_{SB}L_{E1} + \rho_{SP^+}L_{E2})} \tag{6.27}$$

Example 6.4 (a) In an IGBT, the total length of the emitter (L_E) is 6.6 μm. There is no deep P^+ diffusion in the cell structure. What is the shunting resistance R_s if the sheet resistance of the P base is 2500 $\Omega/\square$ and the cell width at right angles to the cross section is 25 μm?

(b) Suppose a P^+ region is built into the structure such that a segment $L_{E1} = 1.2$ μm contains N^+ emitter over P-base diffusion while a segment $L_{E2} = 5.4$ μm contains N^+ emitter over P^+ diffusion. What is the shunting resistance of the composite P and P^+ regions if the sheet resistance of P^+ diffusion is 40 $\Omega/\square$.

(c) Given that half-cell width W_C is 12.5 μm, built-in potential V_{bi} is 0.8 V and current gain of PNP transistor is 0.49, calculate the latching current densities for the two cases as well as their ratio.

(d) If the P^+ diffusion mask is successively altered in such a manner that the segment length L_{E1} is 1.2, 2.5, 3.8, 5.1, and 6.6 μm, calculate the latching current density in each case, and plot the variation of latching current density with segment length L_{E1}.

Figure E6.4.1 shows the IGBT half-cell for this example.

(a) $R_s|_{\text{P-region}} = (1/Z)\rho_{SB}L_E = 1/(25\times10^{-4})\times2500\times6.6\times10^{-4} = 660\ \Omega$.

(b) $R_s|_{\text{P and P}^+} = (1/Z)(\rho_{SB}L_{E1} + \rho_{SB}L_{E2}) = 1/(25\times10^{-4})\times(2500\times1.2\times10^{-4} + 40\times5.4\times10^{-4}) = 128.64\ \Omega$.

(c) $J_{CEL}|_P = V_{bi}/(W_C\alpha_{PNP}\rho_{SB}L_E) = 0.8/(12.5\times10^{-4}\times0.49\times2500\times6.6\times10^{-4}) = 791.589$ A. $J_{CEL}|_{\text{P and P}^+} = V_{bi}/\{W_C\alpha_{PNP}(\rho_{SB}L_{E1} + \rho_{SP^+}L_{E2})\} = 0.8/\{12.5\times10^{-4}\times0.49\,(2500\times1.2\times10^{-4} + 40\times5.4\times10^{-4})\} = 4061.326$ A.

$$\frac{J_{CEL}|_{\text{P and P}^+}}{J_{CEL}|_P} = \frac{4061.326}{791.589} = 5.1306$$

(d) We may write

$$J_{CEL}|_{\text{P and P}^+} = \frac{K}{\rho_{SB}L_{E1} + \rho_{SP^+}L_{E2}}$$

where $K = V_{bi}/(W_C\alpha_{PNP}) = 0.8/(12.5\times10^{-4}\times0.49) = 1306.12$ V/cm. Calculations are shown in the following table (Table E6.4.1).

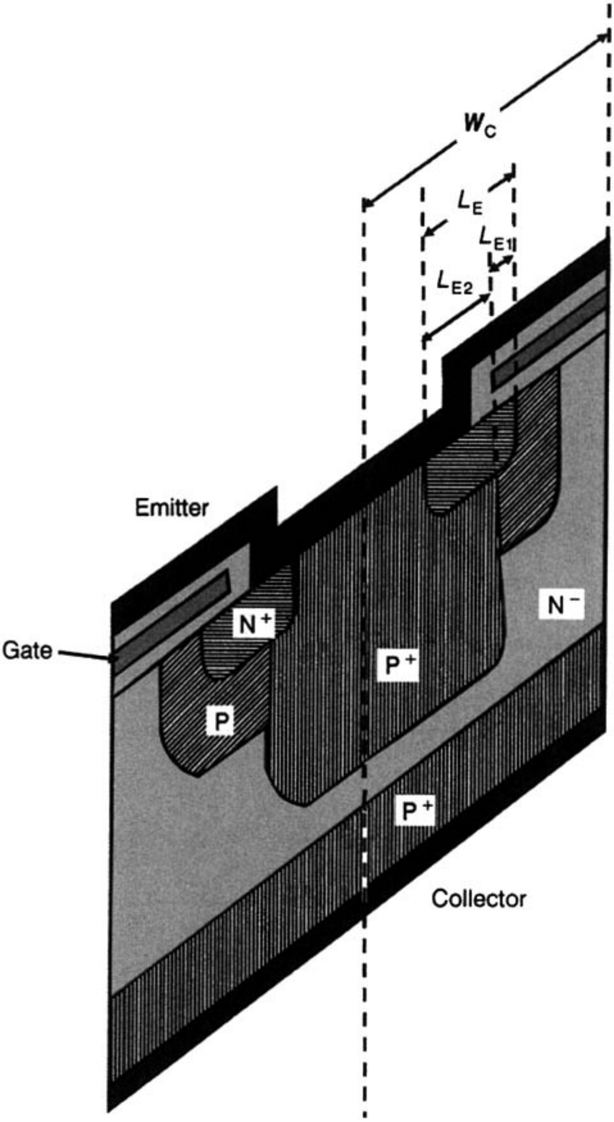

W_C
L_E
L_{E1}
L_{E2}
Emitter
Gate
N+
P
P+
N−
P+
Collector

Figure E6.4.1 IGBT for Example 6.4.

Table E6.4.1 Calculations for Example 6.4

Serial No.	Segment Length L_{E1} (cm)	Segment Length L_{E2}(cm)	$\rho_{SB}L_{E1} + \rho_{SP^+}L_{E2}$	J_{CEL} (A/cm^2)
1	1.2×10^{-4}	5.4×10^{-4}	0.3216	4061.32
2	2.5×10^{-4}	4.1×10^{-4}	0.6414	2036.36
3	3.8×10^{-4}	2.8×10^{-4}	0.9612	1358.84
4	5.1×10^{-4}	1.5×10^{-4}	1.281	1019.61
5	6.6×10^{-4}	0	1.65	791.588

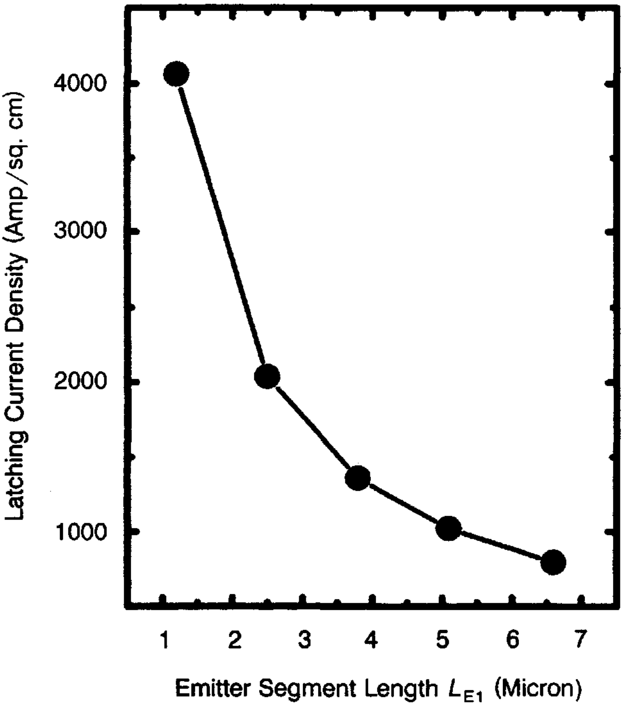


Figure E6.4.2 Plot of latchup current density versus emitter segment length L_{E1}.

Fig. E6.4.2 shows the variation of latching current density with emitter segment length.

---

2. Introduction of a Shallow P^+ Diffusion in the IGBT Cell. Alignment of the deep P^+ diffusion with respect to the polySi gate edge becomes critical for the shorter emitter lengths L_{E1} and therefore imposes a constraint on the smallest achievable length L_{E1}. Implanting a shallow P^+ region [5] within the IGBT cell structure (Fig. 6.7) has solved this problem. This shallow P^+ region is self-aligned to the polySi gate edge. Also, its depth is more than that of the N^+ emitter. These requirements pose practical problems and complicate the processing. For producing a deep P^+ region at a depth of 1 μm, the necessary boron implant energy is 450 keV. At this high-energy value, the polySi gate and the gate oxide may fail to mask the implantation so that the IGBT structure may be disrupted. Furthermore, to create a low sheet resistance region, the implantation dose must be very high. The compensation of N^+ emitter doping by P^+ implant will increase the resistance of N^+ region and hence the forward voltage of IGBT. Decreasing the emitter depth to 0.2 μm has circumvented the difficulties produced by

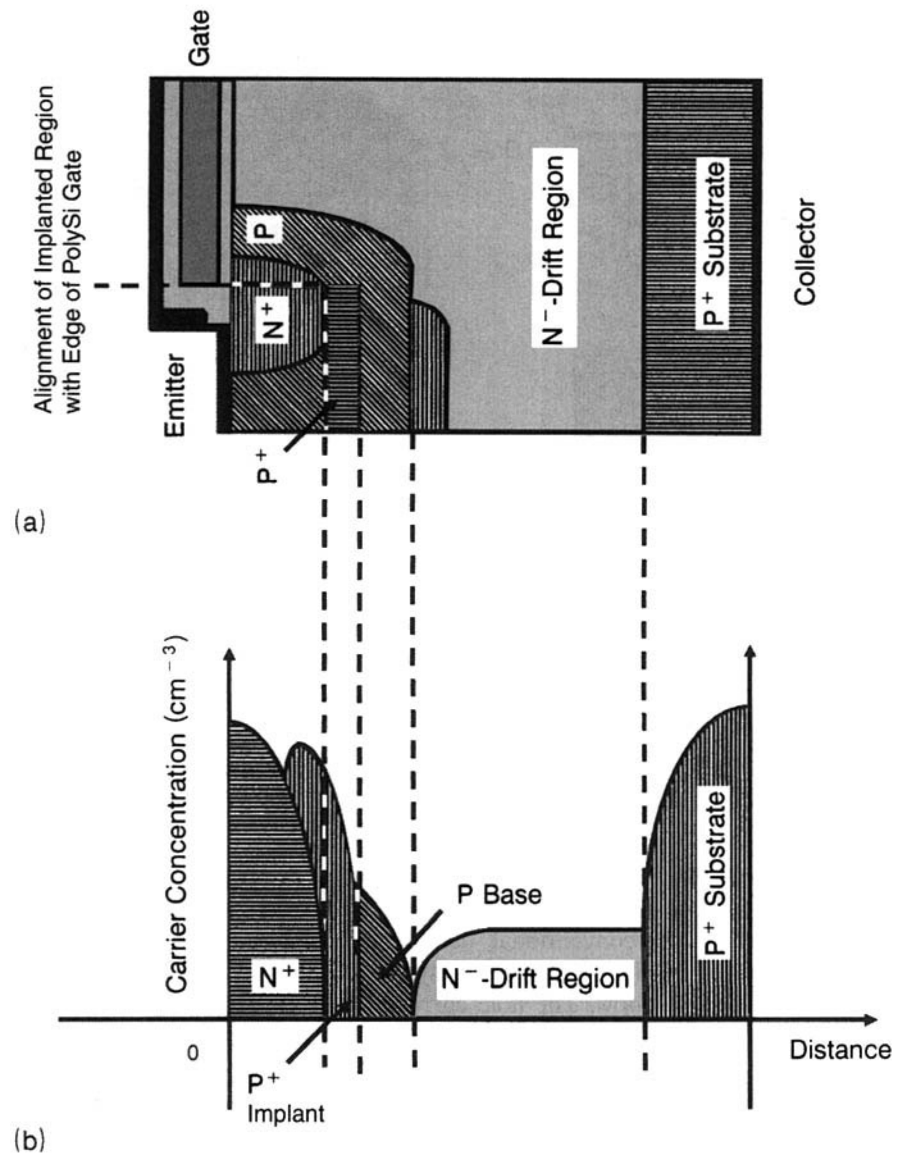

Gate
Alignment of Implanted Region
with Edge of PolySi Gate
Emitter
P
N+
P+
N−-Drift Region
P+ Substrate
Collector
(a)
Carrier Concentration (cm−3)
N+
P Base
N−-Drift Region
P+ Substrate
0
Distance
P+
Implant
(b)

Figure 6.7 (a) Incorporation of shallow P$^+$ implanted region in the IGBT cell structure. (b) Doping profile of IGBT cell with shallow P$^+$ implant.

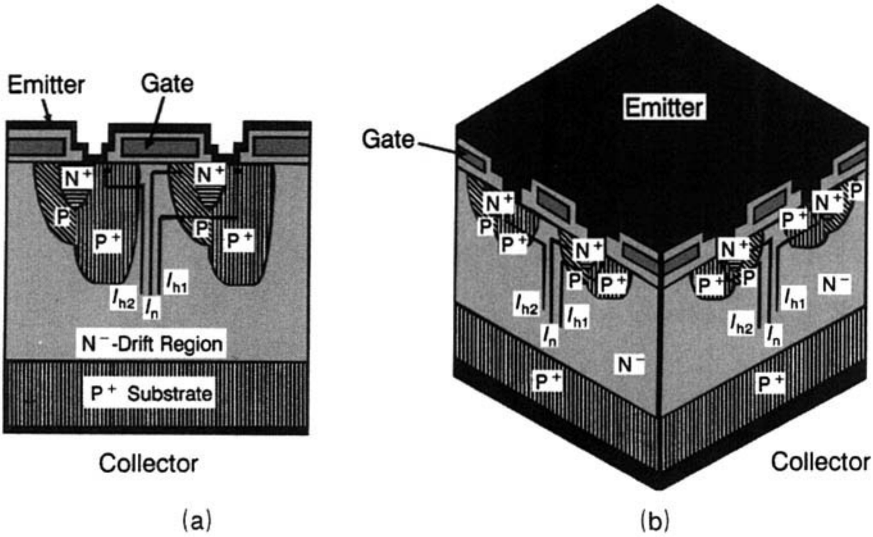


Figure 6.8 Minority-carrier bypass IGBT cell design: (a) Cross section and (b) three-dimensional view.

high-energy and high-dose implantation. When emitter depth is small, implantation energy of 120 keV and dose of 1×10^{14} cm^{-2} is adequate for creating shallow P^{+} region. Double the current density is then required before the device latches.

3. Minority-Carrier By-Pass Cell Design. The conventional IGBT cell design contains two N^{+} emitter regions, one on each side of the polySi gate window. Unlike the conventional design, the minority-carrier bypass [6] design illustrated in Fig. 6.8 contains only one N^{+} emitter region on one side of the polySi gate window. The hole current branches into two paths. In one path, marked I_{h1}, the hole current flows through the P-base region under the N^{+} emitter similar to the conventional IGBT. In the second path marked I_{h2}, the hole current flows via deep P^{+} region without flowing under the N^{+} emitter. Thus the magnitude of hole current traversing underneath the N^{+} emitter is reduced to 50% as compared to normal IGBT operation. Naturally therefore, the latching current density is augmented twofold. The obvious disadvantage of this structure is that the MOSFET channel density is also halved, thereby increasing the forward drop. To overcome this shortcoming of the minority-carrier bypass design, it has been implemented by periodically interrupting the N^{+} emitter along the cell length at right angles to the cross section. The degree of bypassing is traded off with the rise in ON-state voltage drop.

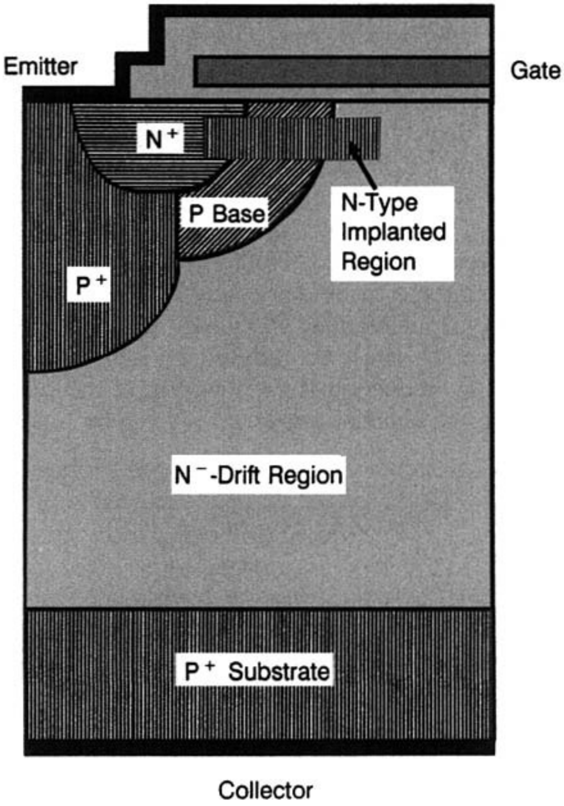


Figure 6.9 Counterdoping of IGBT channel region by N-type ion implantation.

4. Counterdoping of Channel by N-Type Ion Implantation. Using a high dose ion implantation for boron doping decreases the sheet resistance of P base. Then for obtaining an acceptably low threshold voltage, an additional shallow N-type ion implantation, preferably arsenic, is preferred to compensate for the higher boron doping at the surface (Fig. 6.9). This counterdoping [7] is necessary in the channel region but can be done over the full cell area, avoiding the additional photolithographic step. Precise knowledge of boron diffusion profile is essential for carrying out this counterdoping because an excessive doping can initiate spurious turn-on of IGBT without any gate bias. The counterdoping technique is fraught with the disadvantage that the total dopant concentration in the channel region is the sum total of P-base doping concentration and N-type counterdoping concentration. Because the channel mobility depends on the aggregate dopant density in the inversion layer, the resulting degradation of channel mobility by counterdoping increases the forward voltage drop.

5. Gate Oxide Thickness Reduction. From Eq. (3.14), the threshold voltage V_{Th} is approximately proportional to the thickness of gate oxide (t_{ox}) and to the square root of doping concentration of P base (N_{A}), that is, $V_{\mathrm{Th}} \propto t_{\mathrm{ox}}\sqrt{N_{\mathrm{A}}}$. This dependence of V_{Th} allows us to increase the N_{A} and thereby the latching current density [8, 9]. Simultaneously, the threshold voltage is kept low by decreasing t_{ox} proportionately. For example, for ensuring that latching takes place at a higher current density, N_{A} can be raised fourfold. But since $V_{\mathrm{Th}} \propto \sqrt{N_{\mathrm{A}}}$, the threshold voltage will be doubled by this alteration of N_{A}. As a high threshold voltage is undesirable, this increase of threshold voltage has to be counteracted by decreasing t_{ox}. To estimate by what ratio t_{ox} must be reduced for maintaining the same threshold voltage, we observe that $V_{\mathrm{Th}} \propto t_{\mathrm{ox}}$. Hence, t_{ox} must be reduced by half for this purpose. The added advantage of t_{ox} reduction is the lowering of the required gate drive voltage V_{GE} because the channel resistance is given by Eq. (3.23):

$$R_{\mathrm{CH}} = \frac{L_{\mathrm{CH}} t_{\mathrm{ox}}}{Z \mu_{\mathrm{ns}} \varepsilon_0 \varepsilon_{\mathrm{ox}} (V_{\mathrm{GE}} - V_{\mathrm{Th}})} \tag{6.28}$$

where the substitution $C_{\mathrm{ox}} = \varepsilon_0 \varepsilon_{\mathrm{ox}} / t_{\mathrm{ox}}$ has been made; ε_0 is the free-space permittivity and $\varepsilon_{\mathrm{ox}}$ is the dielectric constant of $\mathrm{SiO_2}$. The remaining symbols are the same as defined for Eq. (3.23). Hence if t_{ox} is decreased, a smaller V_{GE} value can be used for the same R_{CH}.

Example 6.5 (a) For latching immunization of power IGBT, the P-base concentration N_{A} is raised. But this profile modification increases the threshold voltage V_{Th} of IGBT to unacceptable magnitudes. If the gate oxide thickness t_{ox} is reduced in the correct proportion to maintain the threshold voltage constant, show that the latching current density J_{L} is approximately inversely proportional to the square of gate oxide thickness. (b) Verify the truth of this relationship for $N_{\mathrm{A}} = 1 \times 10^{17}\ \mathrm{cm^{-3}}$, $t_{\mathrm{ox}} = 1000$ Å; and $N_{\mathrm{A}} = 1.414 \times 10^{17}\ \mathrm{cm^{-3}}$, $t_{\mathrm{ox}} = 500$ Å.

(a) Assuming a clean fabrication process of IGBT, the interface state density Q_{ss} is very low. Neglecting also the contribution from the work function difference term, the threshold voltage equation (3.14) reduces to the form

$$V_{\mathrm{Th}} \cong (2kT/q)\ln(N_{\mathrm{A}}/n_{\mathrm{i}}) + \sqrt{\{4\varepsilon_0\varepsilon_{\mathrm{s}}kTN_{\mathrm{A}}\ln(N_{\mathrm{A}}/n_{\mathrm{i}})\}}\,/(\varepsilon_0\varepsilon_{\mathrm{ox}}/t_{\mathrm{ox}}) \tag{E6.5.1}$$

From this equation it is evident that for a constant V_{Th} value, N_{A}, and t_{ox} can be mutually adjusted as

$$N_{\mathrm{A}} \propto 1/t_{\mathrm{ox}}^2 \tag{E6.5.2}$$

Since the lateral P-base resistance R_{p} varies inversely as the P-base dopant concentration, we have

$$R_{\mathrm{p}} \propto 1/N_{\mathrm{A}} \propto t_{\mathrm{ox}}^2 \tag{E6.5.3}$$

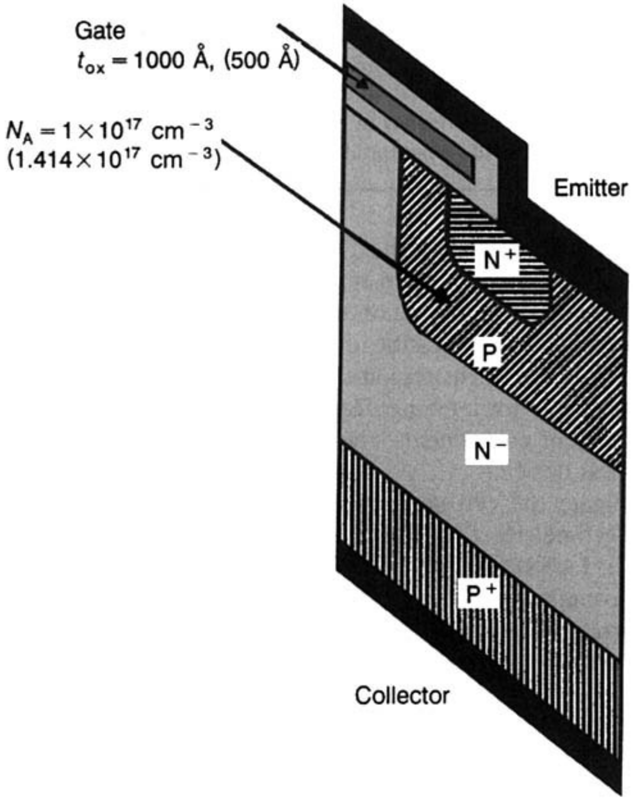

Gate
$t_{ox} = 1000$ Å, (500 Å)
$N_A = 1 \times 10^{17}$ cm^{-3}
(1.414×10^{17} cm^{-3})
Emitter
N+
P
N−
P+
Collector

Figure E6.5.1 IGBT for Example 6.5.

Then from the latching current expression, Eq. (6. 6), we obtain the latching current density

$$J_L \propto 1/R_p \propto 1/t_{ox}^2 \tag{E6.5.4}$$

Thus for a given threshold voltage, the latching current density bears an inverse relationship to the square of gate oxide thickness.

(b) The IGBT half-cell is shown in Fig. E6.5.1. For $N_A = 1 \times 10^{17}$ cm^{-3} and $t_{ox} = 1000$ Å, we have $V_{Th} \cong (2 \times 0.0259)\ln(1 \times 10^{17}/1.45 \times 10^{10}) + \{4 \times 8.854 \times 10^{-14} \times 11.9 \times 1.38 \times 10^{-23} \times 1 \times 10^{17} \times \ln(1 \times 10^{17}/1.45 \times 10^{10})\}^{1/2}/(8.854 \times 10^{-14} \times 3.9/1000 \times 10^{-8}) = 0.81567 + 0.27714 = 1.0928$ V.

Similarly, for $N_A = 1.414 \times 10^{17}$ cm^{-3} and $t_{ox} = 500$ Å, we get $V_{Th} \cong (2 \times 0.0259)\ln(1.414 \times 10^{17}/1.45 \times 10^{10}) + \{4 \times 8.854 \times 10^{-14} \times 11.9 \times 1.38 \times 10^{-23} \times 1.414 \times 10^{17} \times \ln(1.414 \times 10^{17}/1.45 \times 10^{10})\}^{1/2}/(8.854 \times 10^{-14} \times 3.9/\ 500 \times 10^{-8}) = 0.8336 + 0.166579 = 1.000179$ V.

Thus the threshold voltages in the two cases are roughly equal (1.0928 V and 1.000179 V). But the ratio of gate oxide thickness for these cases is = 500 Å/1000 Å = 1/2. So from Eq. (E6.5.4), maintaining the same threshold voltage, the latching current density for the 500 Å-thick gate oxide IGBT will be $1/(1/2)^2 = 4$ times that of the IGBT, having 1000-Å gate oxide thickness. Physically, this happens because when the gate oxide thickness is halved, the P-base doping must be increased to keep the threshold voltage the same. But P-base doping affects the resistance of P-base and therefore the critical current density at which the voltage drop due to hole flow below the emitter equals the built-in potential of the N^+-emitter/P-base junction.

---

6 Choice of IGBT Cell Design. The geometry of IGBT cell determines the subdivision of collector–emitter current I_{CE} into two components: (i) hole current I_h flowing from the region under the polySi gate into the P base and causing latching of the parasitic thyristor and (ii) hole current I_h' flowing directly across the N^--drift region through the P base to the emitter contact and not responsible for latching. Depending upon the relative proportions of these hole current components, the latching current density varies with the cell topological design.

To investigate the cell geometry effect on IGBT latchup, four different IGBT cell geometries (Fig. 6.10)—namely, stripe or linear cell, multiple surface shorts (MSS) cell, square cell, and circular cell—have been compared [10]. These comparisons have revealed that the MSS cell exhibits the highest latching current density. The penalty, however, is paid in terms of the maximum forward voltage drop of the MSS cell, highest amongst these four

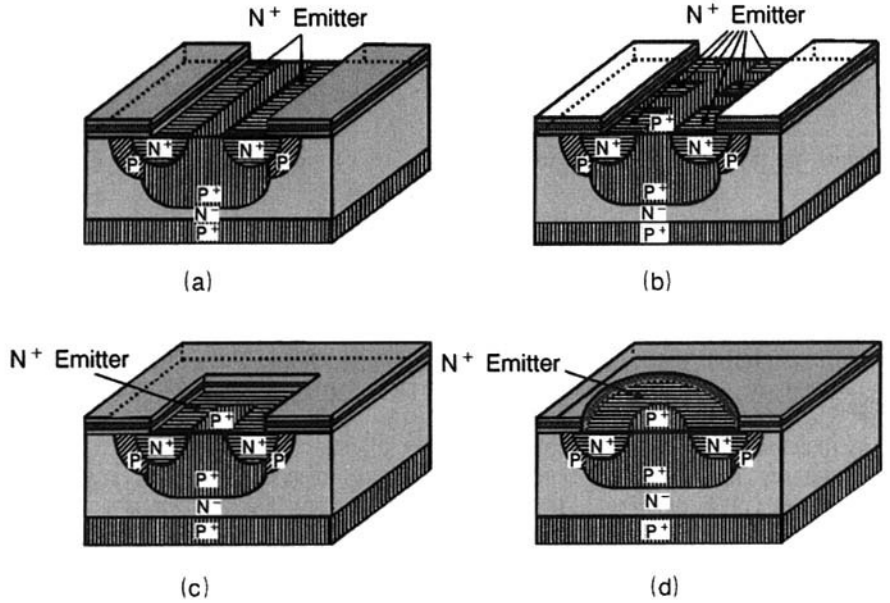


Figure 6.10 Three-dimensional visualization of IGBT cell geometries: (a) Stripe or linear cell, (b) multiple surface shorts (MSS) cell (c) square cell, and (d) circular cell.

cell types. On the decreasing latching current density scale, the MSS cell is followed by the stripe cell which is then followed by the circular cell with the square cell at the bottom. (Example 6.6 performs a comparative analysis of MSS, stripe and circular cells. The latching current densities for these cells are tabulated in this example.). These cells are placed in the reverse order on the decreasing ON-state voltage scale with the square cell at the top and the MSS cell at the bottom. During IGBT design, careful choice of cell geometry calls for compromising between the desired forward voltage drop and latching current density. In the following discussion, equations will be derived for the latching current density expected from different cell geometries, in terms of the shunting resistance for hole flow and the associated structural parameters of IGBT cell.

Stripe or Linear IGBT Cell. Figure 6.11 shows the hole current distribution in the linear cell geometry of IGBT. The hole current component responsible for latching is indicated by the current vectors I_p. If L_G denotes the half-polySi gate width and L_W half-width of the polySi window, we may write

$$I_p = I_{CE}\,\alpha_{PNP}\left(\frac{L_G}{L_G + L_W}\right) \tag{6.29}$$

Referring back to discussions on Eq. (6.26) above, this equation is reached by noting that a fraction $I_{CE}\,\alpha_{PNP}$ of the total collector current I_{CE} of IGBT

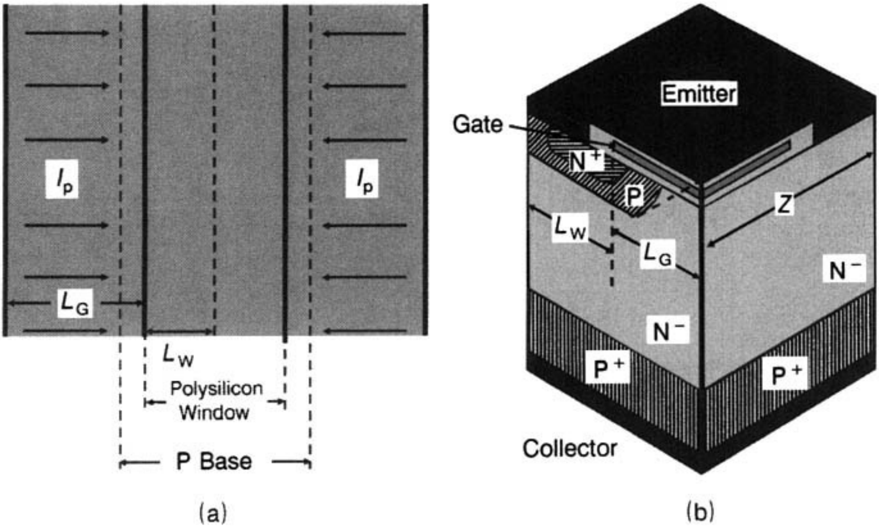


Figure 6.11 (a) Hole current distribution in the linear IGBT cell. (b) Perspective view of the linear cell.

reaches its emitter. Again the collector current I_{CE} pertains to the full area of IGBT unit cell: that is, $2(L_G + L_w) \times Z$, where Z is the width of the linear unit cell. The factor 2 is introduced here because the parameters L_G and L_w correspond to the half-unit cell. Now, imagining the two paths of hole current flow—that is, one moving under the N^+ emitter of IGBT to the emitter contact, and the other flowing vertically upwards from the P^+ substrate through the N^- base directly to the emitter contact without moving below the N^+ emitter—it is evident that the hole current component responsible for latchup originates only from the first path arising from the portion of unit cell having area $= 2L_G \times Z$. Since the remaining hole current component reaching directly the emitter contact is not involved in producing latching, the fraction of hole current responsible for latching is obtained by multiplying $I_{CE}\,\alpha_{PNP}$ by $2L_G \times Z/\{2(L_G + L_W) \times Z\}$, where $\{2(L_G + L_W) \times Z\}$ is the unit cell area. Hence, this hole current is $= I_{CE}\,\alpha_{PNP} \times 2L_G \times Z/\{2(L_G + L_W) \times Z\} = I_{CE}\,\alpha_{PNP}\{L_G/(L_G + L_W)\}$.

The shunting resistance is given by

$$R_s = \frac{\text{P-base resistivity } (\rho_B) \times \text{Distance traveled by current } (L_{E1})}{\text{Cross-sectional area } (A)}$$

$$= \frac{\rho_B L_{E1}}{X_{jP} \times Z} = \frac{\rho_B}{X_{jP}} \times \frac{L_{E1}}{Z} = \rho_{SB} L_{E1}/Z \qquad (6.30)$$

where X_{jP} is the P-base junction depth and $\rho_{sB} = \rho/X_{jP}$ is the sheet resistivity of P base. The condition for latching is that the potential drop across the shunting resistance R_s must become equal to the built-in potential of a diode. Therefore, applying Eqs. (6.29), (6.25), and (6.30), the collector–emitter current at which latching occurs in the linear cell is

$$I_{CEL}\big|_{LIN} = \frac{I_p}{\alpha_{PNP}} \times \left(\frac{L_G + L_w}{L_G}\right) = \frac{V_{bi}/R_s}{\alpha_{PNP}} \times \left(\frac{L_G + L_w}{L_G}\right)$$

$$= \frac{V_{bi}}{R_s\,\alpha_{PNP}} \times \left(\frac{L_G + L_w}{L_G}\right) = \frac{V_{bi} Z}{\rho_{SB} L_{E1}\,\alpha_{PNP}} \times \left(\frac{L_G + L_w}{L_G}\right)$$

$$= \frac{V_{bi} Z}{\alpha_{PNP}\,\rho_{SB} L_{E1}} \times \left(\frac{L_G + L_w}{L_G}\right) \qquad (6.31)$$

and the corresponding collector–emitter current density is

$$J_{CEL}\big|_{LIN} = \frac{I_{CEL}\big|_{LIN}}{Z(L_W + L_G)} = \frac{\dfrac{V_{bi} Z}{\alpha_{PNP}\,\rho_{SB} L_{E1}}\left(\dfrac{L_G + L_w}{L_G}\right)}{Z(L_W + L_G)} = \frac{V_{bi} Z}{\alpha_{PNP}\,\rho_{SB} L_{E1} L_G} \qquad (6.32)$$

since unit cell area $= Z(L_w + L_G)$.

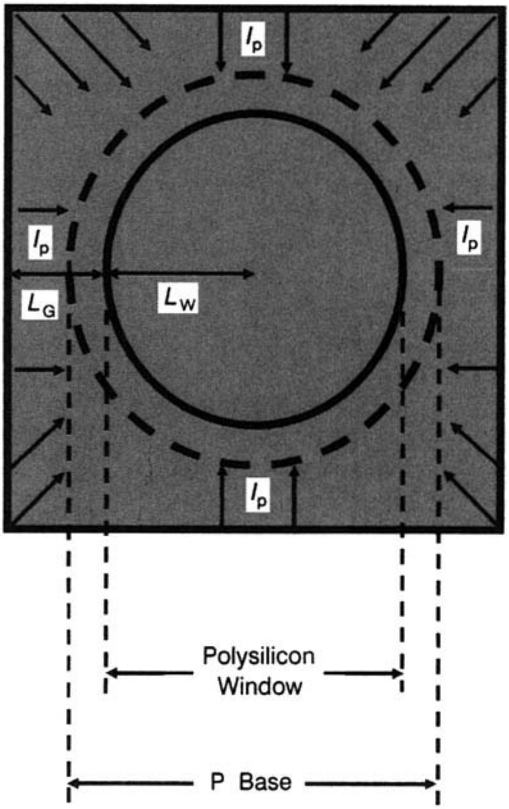


Figure 6.12 Circular IGBT cell showing the hole current distribution.

Circular IGBT Cell. The hole current component I_p converging from all the sides under the polySi gate into the P-base region is indicated by current vectors in Fig. 6.12. In order to determine the portion of IGBT unit cell contributing to the hole current causing latching, we note that the total area of IGBT unit cell equals $\{2(L_W + L_G)\}^2$ and that the area of the window in polysilicon equals πL_w^2. Therefore, area of the annular region from which the hole current producing latchup originates is $4(L_W + L_G)^2 - \pi L_w^2$. This region represents a fraction $\{4(L_W + L_G)^2 - \pi L_w^2\}/\{4(L_W + L_G)^2\}$ of the unit cell area $4(L_W + L_G)^2$. Hence the collector current $I_{CE}\,\alpha_{PNP}$ is multiplied by this fraction to obtain the effective collector current causing latching. The magnitude of the hole current producing latching is

$$I_p = I_{CE}\,\alpha_{PNP}\frac{4(L_W + L_G)^2 - \pi L_W^2}{4(L_W + L_G)^2} \tag{6.33}$$

The expression for shunting resistance is derived by considering a right circular cylinder of radius r and height X_{jP} (depth of P-base region). Let ρ_B

be the bulk resistivity of P base and let ρ_{sB} (its sheet resistivity) $= \rho_{\mathrm{B}}/X_{\mathrm{jP}}$. Then

$$
\begin{aligned}
R_{\mathrm{s}} &= \int_{L_{\mathrm{W}}-L_{\mathrm{E1}}}^{L_{\mathrm{W}}} \frac{\rho_{\mathrm{B}}\,dr}{A} = \int_{L_{\mathrm{W}}-L_{\mathrm{E1}}}^{L_{\mathrm{W}}} \frac{\rho_{\mathrm{B}}\,dr}{2\pi r \times X_{\mathrm{JP}}} = \int_{L_{\mathrm{W}}-L_{\mathrm{E1}}}^{L_{\mathrm{W}}} \frac{\rho_{\mathrm{B}}}{X_{\mathrm{JP}}} \times \frac{dr}{2\pi r} \\
&= \int_{L_{\mathrm{W}}-L_{\mathrm{E1}}}^{L_{\mathrm{W}}} \frac{\rho_{\mathrm{SB}}}{2\pi} \times \frac{dr}{r} = \frac{\rho_{\mathrm{SB}}}{2\pi}[\ln r]_{L_{\mathrm{W}}-L_{\mathrm{E1}}}^{L_{\mathrm{W}}} \\
&= \frac{\rho_{\mathrm{SB}}}{2\pi}\{\ln(L_{\mathrm{W}}) - \ln(L_{\mathrm{W}}-L_{\mathrm{E1}})\} = \frac{\rho_{\mathrm{SB}}}{2\pi}\ln\left(\frac{L_{\mathrm{W}}}{L_{\mathrm{W}}-L_{\mathrm{E1}}}\right) \qquad (6.34)
\end{aligned}
$$

Latching takes place when the voltage drop across R_{s} is sufficient to forward bias the N$^+$-emitter/P-base diode. So, from Eqs. (6.33), (6.25), and (6.34), the collector–emitter current, which starts latching in the circular cell, is

$$
\begin{aligned}
I_{\mathrm{CEI}}|_{\mathrm{CIR}} &= \frac{I_{\mathrm{P}}}{\alpha_{\mathrm{PNP}}} \times \frac{4(L_{\mathrm{W}}+L_{\mathrm{G}})^2}{4(L_{\mathrm{W}}+L_{\mathrm{G}})^2 - \pi L_{\mathrm{W}}^2} \\
&= \frac{V_{\mathrm{bi}}/R_{\mathrm{s}}}{\alpha_{\mathrm{PNP}}} \times \frac{4(L_{\mathrm{W}}+L_{\mathrm{G}})^2}{4(L_{\mathrm{W}}+L_{\mathrm{G}})^2 - \pi L_{\mathrm{W}}^2} \\
&= \frac{V_{\mathrm{bi}}}{\alpha_{\mathrm{PNP}}\,\rho_{\mathrm{SB}}} \frac{8\pi(L_{\mathrm{G}}+L_{\mathrm{W}})^2}{\{4(L_{\mathrm{G}}+L_{\mathrm{W}})^2 - \pi L_{\mathrm{W}}^2\}\ln\{L_{\mathrm{W}}/(L_{\mathrm{W}}-L_{\mathrm{E1}})\}} \qquad (6.35)
\end{aligned}
$$

Then the latching current density for circular cell design is

$$
\begin{aligned}
J_{\mathrm{CEI}}|_{\mathrm{CIR}} &= \frac{I_{\mathrm{CEI}}|_{\mathrm{CIR}}}{4(L_{\mathrm{G}}+L_{\mathrm{W}})^2} \\
&= \frac{\dfrac{V_{\mathrm{bi}}}{\alpha_{\mathrm{PNP}}\,\rho_{\mathrm{SB}}} \dfrac{8\pi(L_{\mathrm{G}}+L_{\mathrm{W}})^2}{\{4(L_{\mathrm{G}}+L_{\mathrm{W}})^2 - \pi L_{\mathrm{W}}^2\}\ln\{L_{\mathrm{W}}/(L_{\mathrm{W}}-L_{\mathrm{E1}})\}}}{4(L_{\mathrm{G}}+L_{\mathrm{W}})^2} \\
&= \frac{V_{\mathrm{bi}}}{\alpha_{\mathrm{PNP}}\,\rho_{\mathrm{SB}}} \frac{2\pi}{\{4(L_{\mathrm{G}}+L_{\mathrm{W}})^2 - \pi L_{\mathrm{W}}^2\}\ln\{L_{\mathrm{W}}/(L_{\mathrm{W}}-L_{\mathrm{E1}})\}} \qquad (6.36)
\end{aligned}
$$

For the circular cell, hole current concentration from the surrounding region into the central polySi window causes it to latch at a lower current density in comparison to the linear cell where the hole current concentrates only from the two sides.

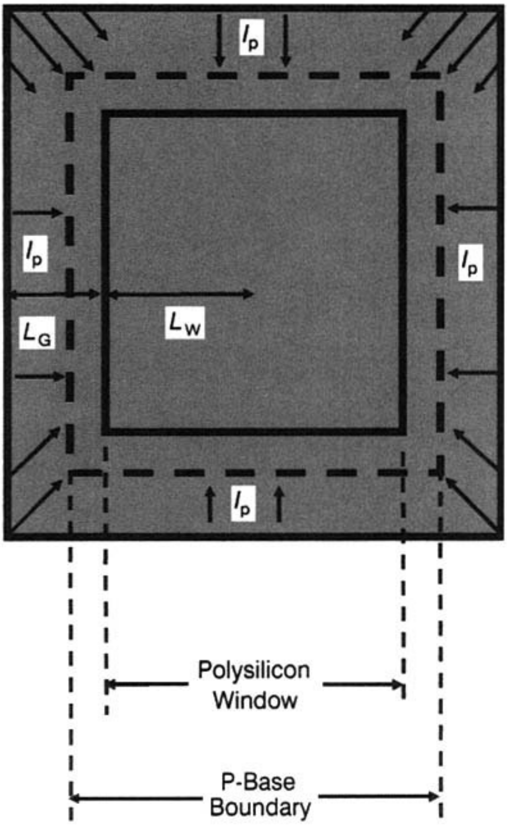


Figure 6.13 Square IGBT cell and the hole current distribution in it.

Square IGBT Cell. Figure 6.13 shows how the hole current converges into the polySi window from the region below the polySi gate on all the sides. This behavior resembles that of the circular cell, but the latching current density for the square cell is lower than that for the circular cell. The reason for this is found by close inspection of the square and circular cell geometries. This reveals that the length of the N^+ emitter stretching over the P-base region (L_{E1}) is larger at the corners of the square cell by $\sqrt{2}$ times than for the circular cell case. This larger emitter length at the corners of the square cell produces additional voltage drop, rendering the square cell prone to latching at a lower current density. Since the circular cell latches at a smaller current density than the linear cell, it is evident that latching current density increases in the order Linear cell > Circular cell > Square cell.

Atomic Lattice Layout (A-L-L) IGBT Cell. As shown in Fig. 6.14, the hole current I_p in the A-L-L IGBT cell geometry spreads outwards from the

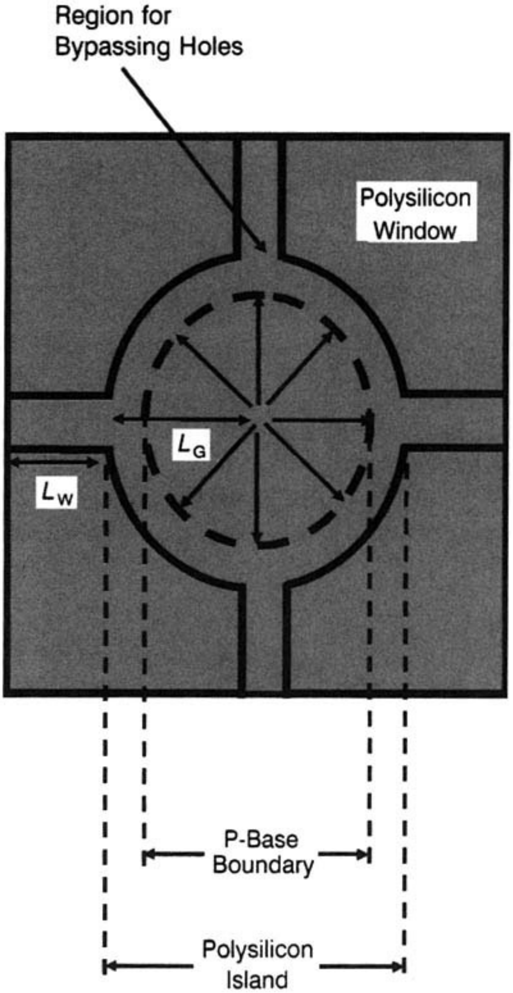


Figure 6.14 Showing the hole current distribution in the atomic-lattice-layout IGBT cell.

region underlying the polySi gate toward the polySi window [11]. At the hole bypass regions, bars of polySi interconnect the circular-shaped polySi pads.

To compare the A-L-L cell design with multiple surface shorts in the linear cell, we find that in the A-L-L cell, the polySi bars do not produce appreciable loss of N^+ emitter area or MOSFET channel width. The reason is that the polySi bars are necessary for fabrication of A-L-L geometry with a single polySi layer. Insertion of the hole bypass regions has a favorable impact on raising the latching current density. Furthermore, the A-L-L cell exhibits superior forward- and reverse-bias safe operating areas (FBSOA and RBSOA) when compared to the linear cell, primarily due to smaller electric field enhancement in this case. The A-L-L cell produces a cylindrical junction rotated opposite to that for a circular cell, creating a saddle junction below the polySi gate. It may also be noted that the electric field enhancement for circular cell is worse than that for the linear cell with rounded corners but smaller than that for the sharp-cornered linear cell.

Latching of the parasitic bipolar thyristor is governed by the total hole current expressed as

$$I_{\mathrm{p}} = I_{\mathrm{CE}}\,\alpha_{\mathrm{PNP}}\frac{\pi L_{\mathrm{G}}^{2}}{4(L_{\mathrm{G}}+L_{\mathrm{W}})^{2}} \tag{6.37}$$

To arrive at this equation, the area below the polySi gate from which the hole current causing latching spreads out is πL_{G}^{2}. Total area of the A-L-L unit cell is $4(L_{\mathrm{G}}+L_{\mathrm{w}})^{2}$. Hence, the hole current producing latchup flows from the fraction $\pi L_{\mathrm{G}}^{2}/4(L_{\mathrm{G}}+L_{\mathrm{w}})^{2}$ of the unit cell area.

In this case, following similar procedure to Eq. (6.34), the shunting resistance R_{s} is

$$\begin{aligned}
R_{\mathrm{s}} &= \int_{L_{\mathrm{G}}}^{L_{\mathrm{G}}+L_{\mathrm{E1}}}\frac{\rho_{\mathrm{B}}\,dr}{A} = \int_{L_{\mathrm{G}}}^{L_{\mathrm{G}}+L_{\mathrm{E1}}}\frac{\rho_{\mathrm{B}}\,dr}{2\pi r\times X_{\mathrm{JP}}} = \int_{L_{\mathrm{G}}}^{L_{\mathrm{G}}+L_{\mathrm{E1}}}\frac{\rho_{\mathrm{B}}}{X_{\mathrm{JP}}}\times\frac{dr}{2\pi r}\\
&= \int_{L_{\mathrm{G}}}^{L_{\mathrm{G}}+L_{\mathrm{E1}}}\frac{\rho_{\mathrm{SB}}}{2\pi}\times\frac{dr}{r} = \frac{\rho_{\mathrm{SB}}}{2\pi}[\ln r]_{L_{\mathrm{G}}}^{L_{\mathrm{G}}+L_{\mathrm{E1}}}\\
&= \frac{\rho_{\mathrm{SB}}}{2\pi}\{\ln(L_{\mathrm{G}}+L_{\mathrm{E1}})-\ln(L_{\mathrm{G}})\} = \frac{\rho_{\mathrm{SB}}}{2\pi}\ln\left(\frac{L_{\mathrm{G}}+L_{\mathrm{E1}}}{L_{\mathrm{G}}}\right)
\end{aligned} \tag{6.38}$$

Then the latching condition is applied. From Eqs. (6.37), (6.25), and (6.38), the collector–emitter current at which latching starts in A-L-L cell is given by

$$\begin{aligned}
I_{\mathrm{CEL}}\big|_{\mathrm{ALL}} &= \frac{I_{\mathrm{p}}}{\alpha_{\mathrm{PNP}}}\times\frac{4(L_{\mathrm{G}}+L_{\mathrm{W}})^{2}}{\pi L_{\mathrm{G}}^{2}} = \frac{V_{\mathrm{bi}}/R_{\mathrm{s}}}{\alpha_{\mathrm{PNP}}}\times\frac{4(L_{\mathrm{G}}+L_{\mathrm{W}})^{2}}{\pi L_{\mathrm{G}}^{2}}\\
&= \frac{V_{\mathrm{bi}}\Big/\left\{\dfrac{\rho_{\mathrm{SB}}}{2\pi}\ln\left(\dfrac{L_{\mathrm{G}}+L_{\mathrm{E1}}}{L_{\mathrm{G}}}\right)\right\}}{\alpha_{\mathrm{PNP}}}\times\frac{4(L_{\mathrm{G}}+L_{\mathrm{W}})^{2}}{\pi L_{\mathrm{G}}^{2}}\\
&= \frac{V_{\mathrm{bi}}}{\alpha_{\mathrm{PNP}}\,\rho_{\mathrm{SB}}}\,\frac{8(L_{\mathrm{W}}+L_{\mathrm{G}})^{2}}{L_{\mathrm{G}}^{2}\ln\{(L_{\mathrm{G}}+{}_{\mathrm{E1}})/L_{\mathrm{G}}\}}
\end{aligned} \tag{6.39}$$

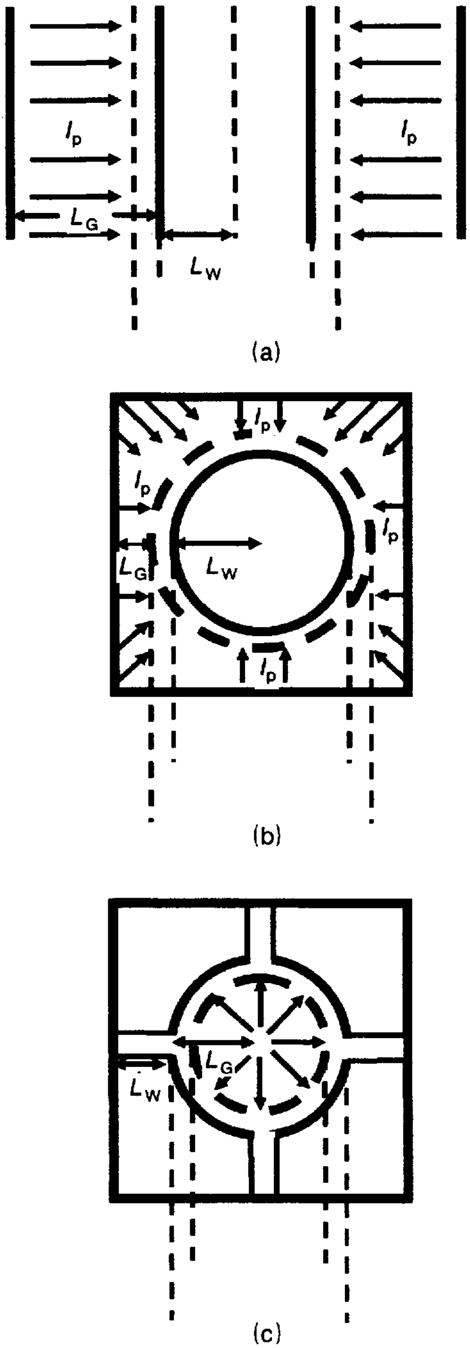

I_p
I_p
L_G
L_W
(a)
I_p
I_p
I_p
I_p
L_G
L_W
(b)
L_G
L_W
(c)

Figure E.6.6.1 Showing L_W, L_G for the three geometries: (a) Stripe cell, (b) circular cell, and (c) atomic lattice layout.

Table E6.6.1 Latching Current Densities J_{CEL} for Various Designs

Serial No.	Cell Topology	Quantity	J_{CEL} (A/cm^2)
1	Linear cell	$\frac{1}{L_{E1}L_G} = 1/(0.9\times10^{-4}\times8\times10^{-4})$ $= 1.389\times10^{7}$	8.4579×10^{3}
2	Circular cell	$\frac{2\pi}{\{4(L_G+L_W)^2-\pi L_W^{\,2}\}\ln\left\{\frac{L_W}{(L_W-L_{E1})}\right\}}$ $= 2\times3.14/[4\{(8+7)\times10^{-4}\}^2$ $-3.14\times(7\times10^{-4})^2] = 4.009\times10^{6}$	2.4414×10^{3}
3	Atomic lattice layout cell	$\frac{2}{L_G^{\,2}\ln\left\{\frac{(L_G+L_{E1})}{L_G}\right\}}$ $= 2/[(8\times10^{-4})^2\ln\{(8+0.9)$ $\times10^{-4}/(8\times10^{-4})\}]$ $= 2.931253\times10^{7}$	17.85×10^{3}

yielding the latching current density

$$J_{\mathrm{CEL}}\big|_{\mathrm{ALL}} = \frac{I_{\mathrm{CEL}}\big|_{\mathrm{ALL}}}{4(L_G+L_W)^2} = \frac{\dfrac{V_{\mathrm{bi}}}{\alpha_{\mathrm{PNP}}\,\rho_{\mathrm{SB}}}\,\dfrac{8(L_W+L_G)^2}{L_G^{\,2}\ln\{(L_G+L_{E1})/L_G\}}}{4(L_G+L_W)^2}$$

$$= \frac{2V_{\mathrm{bi}}}{\alpha_{\mathrm{PNP}}\,\rho_{\mathrm{SB}}\,L_G^{\,2}\ln\{(L_G+L_{E1})/L_G\}} \tag{6.40}$$

Example 6.6 For an IGBT cell having the parameters $L_W = 7\ \mu$m, $L_G = 8\ \mu$m, and $L_{E1} = 0.9\ \mu$m, calculate the latching current densities for linear, circular, and A-L-L designs, given that $\alpha_{\mathrm{PNP}} = 0.52$, $V_{\mathrm{bi}} = 0.95$ V, $W_C = 17\ \mu$m, and $\rho_{\mathrm{SB}} = 3000$ ohm per square ($\Omega/\square$). [ohm per square is the unit of sheet resistivity (symbol ρ_s) defined as the resistance of a square-shaped specimen of the material in the form of a sheet. The relation between sheet resistivity ρ_s and bulk resistivity ρ of the material is $\rho_s = \rho/t$.]

Figure E6.6.1 shows the relevant parameters L_w and L_G for the three geometrical designs. The constant $K = V_{\mathrm{bi}}/(\alpha_{\mathrm{PNP}}\rho_{\mathrm{SB}}) = 0.95/(0.52\times3000) = 6.0897\times10^{-4}$ Volt-ohm per square (V-$\square/\Omega$). The latching current densities J_{CEL} for the various designs are compiled in Table E6.6.1.

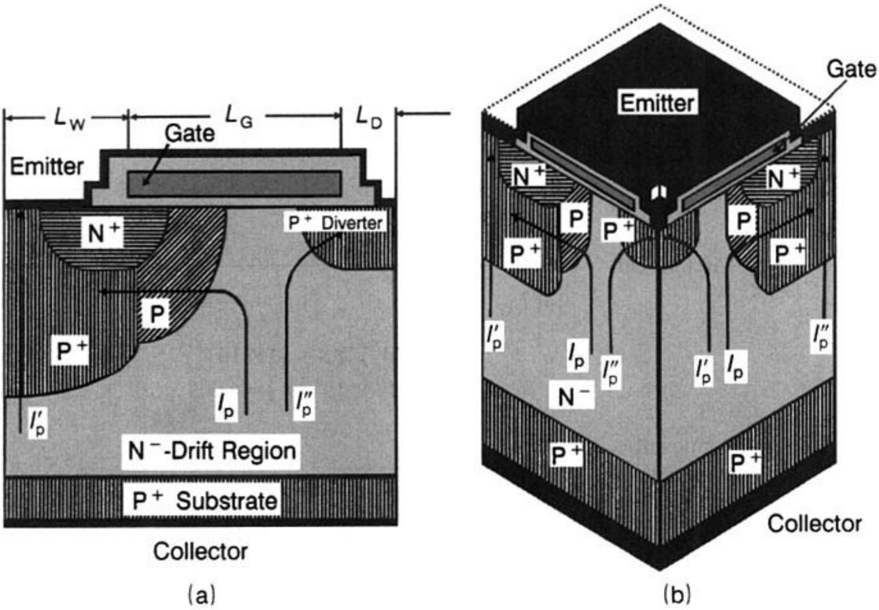


Figure 6.15 Addition of a deep P^+ diverter region in the IGBT cell: (a) Cross-sectional diagram and (b) isometric view.

7. IGBT Cell Structure with Diverter Region (Fig. 6.15). Heretofore, we have seen that the total hole current has two components I_p and I'_p. In the linear IGBT with diverter region [12], the junction between the P^+ diverter region and the N-drift region introduces a third path I''_p (Fig. 6.15). This diverter region is connected to the emitter metal, and therefore P^+ diverter/N^--drift region junction is reverse-biased. The diverter region is incorporated in any of the aforementioned cell geometries. For the linear cell topology, assume that the hole current I_p causing latchup flows from half the area under the polySi gate while the hole current from the remaining area under the polySi gate flows into the diverter region. Then the hole current producing latching of thyristor is

$$I_p = I_{CE}\,\alpha_{PNP}\frac{L_G}{2(L_G + L_W + L_D)} \tag{6.41}$$

where L_D is half-width of the diverter region. It may be noted that the unit cell area from which hole current producing latchup originates is expressed as follows: Area below the polySi gate/2 = $(1/2) \times L_G \times Z$, where Z is the width of the cell. Total unit cell area = $(L_G + L_W + L_D) \times Z$. So, the hole current $I_{CE}\,\alpha_{PNP}$ is multiplied by the ratio $\{(1/2) \times L_G \times Z\}/(L_G + L_W + L_D) \times Z = L_G/\{2(L_G + L_W + L_D)\}$ to calculate the hole current to which latching is ascribed.

The shunt resistance for this cell is the same as for the linear cell given in Eq. (6.30) above:

$$R_s = \frac{\rho_{SB} L_{E1}}{Z} \tag{6.42}$$

Latchup occurs when

$$\begin{aligned} I_{CEL}\big|_{DIV} &= \frac{I_P}{\alpha_{PNP}} \times \frac{2(L_G + L_W + L_D)}{L_G} = \frac{V_{bi}/R_s}{\alpha_{PNP}} \times \frac{2(L_G + L_W + L_D)}{L_G} \\ &= \frac{V_{bi}\Big/\left(\frac{\rho_{SB} L_{E1}}{Z}\right)}{\alpha_{PNP}} \times \frac{2(L_G + L_W + L_D)}{L_G} \\ &= \frac{V_{bi} Z}{\alpha_{PNP}\,\rho_{SB}} \left\{\frac{2(L_G + L_W + L_D)}{L_G L_{E1}}\right\} \end{aligned} \tag{6.43}$$

where Eqs. (6.42), (6.25), and (6.43) have been used.

The collector-emitter latching current density is

$$\begin{aligned} J_{CEL}\big|_{DIV} &= \frac{I_{CEL}\big|_{DIV}}{Z(L_G + L_W + L_D)} = \frac{\frac{V_{bi} Z}{\alpha_{PNP}\,\rho_{SB}} \left\{\frac{2(L_G + L_W + L_D)}{L_G L_{E1}}\right\}}{Z(L_G + L_W + L_D)} \\ &= \frac{2V_{bi}}{\alpha_{PNP}\,\rho_{SB} L_{E1} L_G} \end{aligned} \tag{6.44}$$

since unit cell area $= Z(L_G + L_W + L_D)$. Comparing this equation with that for the simple linear IGBT cell, Eq. (6.32), we find that the linear cell with diverter region doubles the latching current density with respect to the simple linear cell. Over and above, the diverter region serves as a guard ring for the junction between the P-base and N-drift region, improving the forward-biased safe operating area (FBSOA) of IGBT. The disadvantage associated with the diverter region is that by virtue of JFET action, the resistance to electron flow from the channel increases. The forward voltage drop of IGBT undesirably rises. This is experienced particularly for the small separation between the P base and diverter junction and when the junction depth of the diverter equals that of the P base. The voltage drop degradation, can be reduced by using a smaller junction depth for diverter region, along with a larger separation from P base.

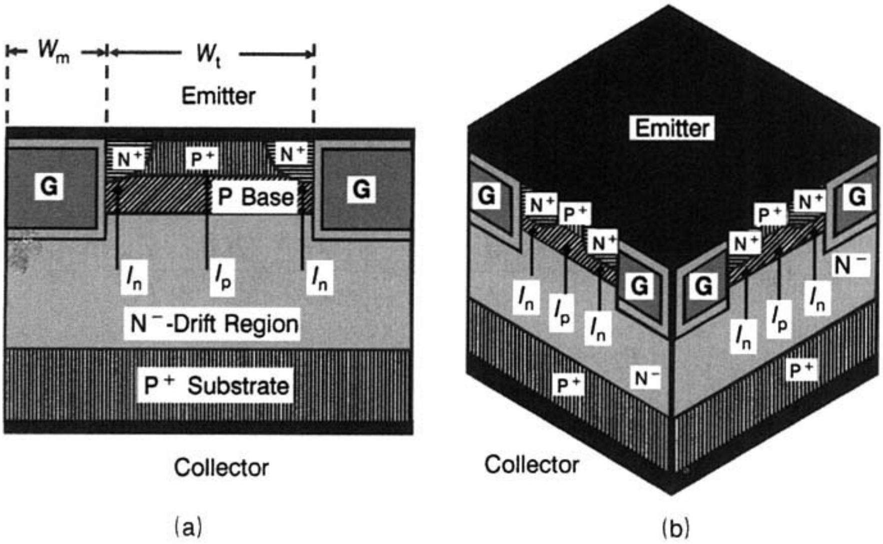


Figure 6.16 Geometrical parameters for analyzing latching current density of trench-gate IGBT: (a) Cross section and (b) isometric drawing.

6.5 LATCHING CURRENT DENSITY OF TRENCH-GATE IGBT (FIG. 6.16)

The trench-gate UMOS IGBT structure provides a significant improvement in the latching current capability in comparison to the DMOS structure. The reason is the difference in hole current flow paths in the two cases. In the DMOS structure, the hole current mainly flows horizontally underneath the N^+ emitter causing a large potential drop, which brings about latchup at lower collector–emitter currents. In the UMOS IGBT (Fig. 6.16), the hole current flows vertically, and the resistance to hole current depends on the depth of the N^+ emitter. By using a shallow P^+ region, as shown in the diagram, this resistance can be appreciably reduced, thereby making the IGBT more latchup-resistant. This P^+ region plays the same role as the deep P^+ well of Fig. 6.6b commonly used for preventing latchup. It reduces the resistance to hole current. However, it has a different approach from the P^+ diverter of Fig. 6.15, the reason being that the latter does not reduce the path resistance to hole current but instead decreases the hole current component flowing into the P-base region.

Latching current density of a trench-gate IGBT is determined by following the same procedure as for DMOS IGBT. As mentioned above, the hole current component I_p responsible for forward biasing the N^+-emitter/P-base junction is governed by the junction depth x_{N^+} of N^+ emitter and the resistivity ρ_{P^+} of P^+ layer . The shunting resistance R_s can be expressed as

$$R_s = \frac{\rho_{\mathrm{P}^+} \times \text{length traversed}}{\text{Area}} = \frac{\rho_{\mathrm{P}^+}}{Z} \frac{2x_{\mathrm{N}^+}}{W_m - 2x_{\mathrm{N}^+}} \tag{6.45}$$

where W_m is the distance between adjacent trenches. Sheet resistivity of P^+ layer is not used here because of the vertical flow of hole current. The condition of latching is given by Eq. (6.25) as

$$I_p R_s = V_{bi} \tag{6.46}$$

where hole current I_p is expressed as

$$I_p = \alpha_{\mathrm{PNP}} Z \frac{W_m + W_t}{2} J_{\mathrm{CE}} \tag{6.47}$$

The latching current density $(J_{\mathrm{CEL}})_{\mathrm{Trench}}$ is written as

$$\begin{aligned}(J_{\mathrm{CEL}})_{\mathrm{Trench}} &= \frac{I_p}{\alpha_{\mathrm{PNP}}} \times \frac{2}{Z(W_m + W_t)} = \frac{V_{bi}/R_s}{\alpha_{\mathrm{PNP}}} \times \frac{2}{Z(W_m + W_1)} \\ &= \frac{V_{bi} \Big/ \left(\dfrac{\rho_{\mathrm{P}^+}}{Z} \dfrac{2x_{\mathrm{N}^+}}{W_m - x_{\mathrm{N}^+}} \right)}{\alpha_{\mathrm{PNP}}} \times \frac{2}{Z(W_m + W_1)} \\ &= \frac{V_{bi}}{\alpha_{\mathrm{PNP}} \rho_{\mathrm{P}^+} x_{\mathrm{N}^+}} \frac{W_m - 2x_{\mathrm{N}^+}}{W_m + W_1}\end{aligned} \tag{6.48}$$

where Eqs. (6.47), (6.46), and (6.45) have been applied.

Now using Eq. (6.27), we find that

$$\begin{aligned}\frac{(J_{\mathrm{CEL}})_{\mathrm{Trench}}}{(J_{\mathrm{CEL}})_{\mathrm{DMOS}}} &= \frac{\dfrac{V_{bi}}{\alpha_{\mathrm{PNP}} \rho_{\mathrm{P}^+} x_{\mathrm{N}^+}} \dfrac{W_m - 2x_{\mathrm{N}^+}}{W_m + W_1}}{\dfrac{V_{bi}}{W_{\mathrm{C}} \alpha_{\mathrm{PNP}} (\rho_{\mathrm{SB}} L_{\mathrm{E1}} + \rho_{\mathrm{SP}^+} L_{\mathrm{E2}})}} \\ &= \frac{\rho_{\mathrm{SB}} L_{\mathrm{E1}} + \rho_{\mathrm{SP}^+} L_{\mathrm{E2}}}{\rho_{\mathrm{P}^+}} \frac{W_m - 2x_{\mathrm{N}^+}}{x_{\mathrm{N}^+}} \frac{(W_{\mathrm{Cell}})_{\mathrm{DMOS}}}{(W_{\mathrm{CEll}})_{\mathrm{Trench}}}\end{aligned} \tag{6.49}$$

where $(W_{\mathrm{Cell}})_{\mathrm{Trench}} = W_m + W_t$ and $(W_{\mathrm{Cell}})_{\mathrm{DMOS}} = W_{\mathrm{C}}$. Besides lower resistance to hole current, the smaller cell size in trench-gate IGBT is responsible for higher latching current density of this IGBT than in the DMOS case.

6.6 SUMMARIZING REMARKS

The curse of latching was a serious menace in the early phase of IGBT development. But with the recent innovations in latchup suppression techniques and introduction of novel designs and structures, a strong immunity against latching has been built up in the IGBT. The latching drawback has been satisfactorily controlled.

REVIEW EXERCISES

6.1 Explain how does the latching of parasitic thyristor in IGBT take place? Distinguish between static and dynamic latching of IGBT.

6.2 Considering a simple IGBT cell structure without deep P^+ diffusion, formulate the dependence of static latching current $I_{L,SS}$ on the current gain of PNP transistor (α_{PNP}), sheet resistance of P base (ρ_{SB}), and emitter length (L_E).

6.3 Explain how the built-in potential of the N^+P diode, current gains of NPN and PNP transistors and P-base resistance are affected by temperature to aggravate the latching problem in IGBTs at high temperatures.

6.4 Derive the equation for the ratio of dynamic latching current to static latching current ($I_{L,DR}/I_{L,SS}$) for an IGBT with a resistive load in a DC circuit. Comment on the degradation of latching current with DC supply voltage used, and the current gain of PNP transistor (α_{PNP}).

6.5 Give arguments for the fact that the reduction in latching current density for an asymmetric IGBT under gate-controlled turn-off with a resistive load in a DC circuit will be worse than for the symmetrical IGBT case.

6.6 Show that the decrease in dynamic latching current for an IGBT with an inductive load in a DC circuit will be larger than for a resistive load.

6.7 Modern IGBTs are fabricated with a deep P^+ region in the center of the DMOS cell. Derive an equation for the latching current density of an IGBT cell in which the P base having a sheet resistance ρ_{SB} extends over an emitter length L_{E1} with a sheet resistance ρ_{SP^+}, and the deep P^+ diffusion stretches over an emitter length L_{E2}.

6.8 How does the introduction of a deep P^+ diffusion in IGBT structure impose a limitation on the smallest possible emitter length L_{E1}. How is the problem overcome by employing a shallow ion-implanted P^+ region. Highlight the practical problems encountered in implementing this technique.

6.9 How does the minority-carrier bypass design reduce the hole current flow underneath the N^+ emitter to one-half that in the common IGBT structure. Mention the effect of this design on the ON-state voltage drop of IGBT.

6.10 How does the counterdoping of channel help in raising the latching current density of IGBT keeping the threshold voltage at an acceptable level. Does this method necessitate an additional photomask? Name the dopant that is preferable for this implantation. Point out the practical difficulties faced and give one disadvantage of this method.

6.11 How does the use of a thick gate oxide aid in achieving a higher latching current density in IGBT. Does a smaller gate oxide thickness allow reduction in gate drive voltage? Why?

6.12 What is the role of IGBT cell topological design in determining the latching current density of the device? Why is the latching current density of circular and square IGBT cells lower than that of the linear cell? Why is the latching current density of square IGBT cell inferior to that of the circular cell? With reference to the comparison of latching current densities of cells, give your remarks about the atomic lattice layout geometry.

6.13 Draw a labeled diagram illustrating the hole current distribution in a linear IGBT cell. Derive an equation for the latching current density of this cell in terms of emitter length extending over P base (L_{E1}), half-polySi gate width (L_G), built-in potential V_{bi} between N^+ emitter and P base, current gain of PNP transistor (α_{PNP}), and sheet resistance of P-base.

6.14 Show diagrammatically the hole current distributions in circular, square, and atomic-lattice-layout cells. Explain how does the effect of hole concentration into the polySi window exhibit reversal in the case of A-L-L geometry?

6.15 For the circular and A-L-L IGBT cell designs, write the equations for (i) the hole current component I_h responsible for latchup of parasitic thyristor, (ii) the shunting resistance R_s, and (iii) the latching current density. Explain the meanings of the symbols used.

6.16 Can the diverter region be incorporated into any of the IGBT cell topologies? Analyze the impact of the diverter region on latching in linear cell geometry, assuming that the hole current component causing latchup flows from 50% of the area under the gate electrode.

6.17 How does the use of a diverter region help in improving the forward-biased safe operating area of IGBT? Point out its influence on the forward drop of IGBT.

6.18 Name two factors accounting for the higher latching current density of trench-gate IGBT as compared to the DMOS structure.

6.19 The current gain of PNP transistor in an IGBT is 0.4. The doping concentration of P base is 8.5×10^{16} cm^{-3}. Junction depth of P base is 3 μm. Emitter diffusion is carried out through a 5-μm window up to a depth of 0.9 μm. The experimentally measured latching current density of the IGBT is 1000 A/cm^2. If now the P base concentration is raised to 2×10^{17} cm^{-3} and the emitter diffusion window is taken as 3 μm, what will be the new latching current density?

6.20 Determine the ratio (dynamic latching current/static latching current) for an NPT-IGBT in a DC circuit with resistive load and a supply voltage of 400 V. The N^--base doping concentration is 5×10^{14}cm^{-3}, and its thickness is 125 μm. Ambipolar lifetime in N^- base is 7 μsec. Electron and hole diffusion constants are 35 and 12 cm^2/sec, respectively.

6.21 A PT-IGBT has a 35-μm-thick N^--base region and 5-μm-thick N-buffer layer. If the ambipolar diffusion length is 140 μm, calculate the ratio of dynamic to static latching current for this PT-IGBT. What will happen when the diffusion length becomes 40 μm?

6.22 For latchup prevention in IGBT, the P-base concentration is increased. If the gate oxide thickness is decreased to keep the threshold voltage constant, the latching current density varies inversely with the (gate oxide thickness)^2. Show that this relation is valid for P-base dopant density $N_A = 1 \times 10^{17}$ cm^{-3}, and gate oxide thickness $t_{ox} = 900$ Å; as well as for $N_A = 1.414 \times 10^{17}$ cm^{-3}and $t_{ox} = 450$ Å.

REFERENCES

1. B. J. Baliga, *Revolutionary Innovations in Power Discrete Devices*, IEEE Tech. Digest IEDM, 1986, pp. 102–105.
2. B. J.Baliga, Switching Speed Enhancement in Insulated Gate Transistors by Electron Irradiation, *IEEE Trans. Electron Devices*, Vol. ED-31, No. 12, December 1984, pp. 1790–1795.
3. A. M. Goodman, J. P. Russell, L. A. Goodman, J. C. Nuese, and J. M. Neilson, Improved COMFETs with Fast Switching Speed and High Current Capability, in *International Electron Devices Meeting, IEDM Technical Digest, IEEE*, New York, Abstract 4.3, 1983, pp. 79–82.
4. B. J. Baliga, M. S. Adler, P. V. Gray, and R. P. Love, Suppressing Latch-up in Insulated Gate Transistors, *IEEE Electron Device Lett.*, Vol. EDL-5, No. 8, August 1984, pp. 323–325.
5. S. Eranen and M. Blomberg, The Vertical IGBT with an Implanted Buried Layer, in *IEEE International Symposium Power Semiconductor Devices & ICs (ISPSD91)*, IEEE, New York, 1991, p. 211.
6. A. Nakagawa, H. Ohashi, M. Kurata, H. Yamaguchi, and K. Watanabe, Non-latch-up, 1200 V Bipolar Mode MOSFET with Large SOA, in *International Electron Devices Meeting, IEDM Technical Digest, IEEE*, New York, Abstract 16.8, 1984, pp. 860–861.
7. T. P. Chow, B. J. Baliga and D. N. Pattanayak, Counter-Doping of MOS Channel (CDC)—A New Technique of Improving Suppression of Latching in Insulated Gate Bipolar Transistors, *IEEE Electron Device Lett.*, IEEE, New York, Vol. EDL-9, No. 1, January 1988, pp. 29–31.
8. T. P. Chow and B. J. Baliga, The Effect of MOS Channel Length on the Performance of Insulated Gate Transistors, *IEEE Electron Device Lett.*, IEEE, New York, Vol. EDL-6, No. 8, August 1985, pp. 413–415.
9. T. P. Chow and B. J. Baliga, The Effect of Channel Length and Gate Oxide Thickness on the Performance of Insulated Gate Transistors, *IEEE Trans.Electron Devices*, Vol. ED-32, No. 11, November 1985, p. 2554, Abstract VIB-4, 43rd Annual Device Research Conference, June 17-19, 1985, Boulder, CO.
10. H. Yilmaz, Cell Geometry Effect on IGBT Latch-Up, *IEEE Electron Device Lett.*, Vol. EDL-6, No. 8, August 1985, pp. 419–421.
11. B. J. Baliga, S. R. Chang, P. V. Gray, and T. P. Chow, New Cell Designs for Improving IGBT Safe Operating Area, in *International Electron Devices Meeting, IEDM Technical Dig.*, IEEE, New York, Abstract 34.5, 1988, pp. 809–812.
12. N. Thapar and B. J. Baliga, A New IGBT Structure with Wider Safe Operating Area (SOA), in *IEEE Symposium on Power Semiconductor Devices & ICs (ISPSD94)*, IEEE., New York, Abstract 4.3, 1994, p.177.

APPENDIX 6.1: EQUATION (6.15)

$$\frac{1-\alpha_{\mathrm{PNP,DR}}}{1-\alpha_{\mathrm{PNP,SS}}}=\frac{1-\dfrac{1}{\cosh(W_{\mathrm{DR}}/L_{\mathrm{a}})}}{1-\dfrac{1}{\cosh(2d/L_{\mathrm{a}})}}=\frac{\dfrac{\cosh(W_{\mathrm{DR}}/L_{\mathrm{a}})-1}{\cosh(W_{\mathrm{DR}}/L_{\mathrm{a}})}}{\dfrac{\cosh(2d/L_{\mathrm{a}})-1}{\cosh(2d/L_{\mathrm{a}})}}$$

$$=\left\{\frac{\cosh(W_{\mathrm{DR}}/L_{\mathrm{a}})-1}{\cosh(2d/L_{\mathrm{a}})-1}\right\}\times\left\{\frac{\cosh(2d/L_{\mathrm{a}})}{\cosh(W_{\mathrm{DR}}/L_{\mathrm{a}})}\right\} \qquad (\mathrm{A6.1.1})$$

First factor of Eq. (A6.1.1) is

$$\left\{\frac{\cosh(W_{\mathrm{DR}}/L_{\mathrm{a}})-1}{\cosh(2d/L_{\mathrm{a}})-1}\right\}=\frac{\dfrac{\exp(W_{\mathrm{DR}}/L_{\mathrm{a}})+\exp(-W_{\mathrm{DR}}/L_{\mathrm{a}})}{2}-1}{\dfrac{\exp(2d/L_{\mathrm{a}})+\exp(-2d/L_{\mathrm{a}})}{2}-1}$$

$$=\frac{\exp(W_{\mathrm{DR}}/L_{\mathrm{a}})+\exp(-W_{\mathrm{DR}}/L_{\mathrm{a}})-2}{\exp(2d/L_{\mathrm{a}})+\exp(-2d/L_{\mathrm{a}})-2}$$

$$\cong\frac{\exp(W_{\mathrm{DR}}/L_{\mathrm{a}})-2}{\exp(2d/L_{\mathrm{a}})-2} \qquad (\mathrm{A6.1.2})$$

Equation (A6.1.2) is obtained because $\exp(-W_{\mathrm{DR}}/L_{\mathrm{a}})$ is much smaller than $(W_{\mathrm{DR}}/L_{\mathrm{a}})$, and likewise $\exp(-2d/L_{\mathrm{a}})$ is $\ll \exp(2d/L_{\mathrm{a}})$ as explained in Section 6.3.1 with reference to Eq. (6.12). Substituting for W_{DR} from Eq. (6.11) into Eq. (A6.1.2), we get

$$\text{First factor}=\frac{\exp\left(\dfrac{2d-\sqrt{\dfrac{2\varepsilon_0\varepsilon_{\mathrm{s}}V_1}{qN_{\mathrm{D}}}}}{L_{\mathrm{a}}}\right)-2}{\exp(2d/L_{\mathrm{a}})-2} \qquad (\mathrm{A6.1.3})$$

Now second factor of Eq. (A6.1.1) is

$$\left\{\frac{\cosh(2d/L_{\mathrm{a}})}{\cosh(W_{\mathrm{DR}}/L_{\mathrm{a}})}\right\}=\frac{\dfrac{\exp(2d/L_{\mathrm{a}})+\exp(-2d/L_{\mathrm{a}})}{2}}{\dfrac{\exp(W_{\mathrm{DR}}/L_{\mathrm{a}})+\exp(-W_{\mathrm{DR}}/L_{\mathrm{a}})}{2}}$$

$$=\frac{\exp(2d/L_{\mathrm{a}})}{\exp(W_{\mathrm{DR}}/L_{\mathrm{a}})} \qquad (\mathrm{A6.1.4})$$

neglecting $\exp(-2d/L_a)$ in comparison to $\exp(2d/L_a)$ and $\exp(-W_{DR}/L_a)$ as compared to $\exp(W_{DR}/L_a)$. Putting the value of W_{DR} from Eq. (6.11) into Eq. (A6.1.4), we have

$$\text{Second factor} = \frac{\exp(2d/L_a)}{\exp\left(\dfrac{2d - \sqrt{\dfrac{2\varepsilon_0\varepsilon_s V_1}{qN_D}}}{L_a}\right)} \tag{A6.1.5}$$

Combining the first and second factors of Eq. (A6.1.1) from Eqs. (A6.1.3) and (A6.1.5), we obtain

$$\frac{1-\alpha_{PNP,DR}}{1-\alpha_{PNP,SS}} = \frac{\exp\left(\dfrac{2d}{L_a}\right)}{\exp\left(\dfrac{2d}{L_a}\right)-2} \frac{\exp\left(\dfrac{2d - \sqrt{\dfrac{2\varepsilon_0\varepsilon_s V_1}{qN_D}}}{L_a}\right)-2}{\exp\left(\dfrac{2d - \sqrt{\dfrac{2\varepsilon_0\varepsilon_s V_1}{qN_D}}}{L_a}\right)}$$

$$= \frac{\exp\left(\dfrac{2d}{L_a}\right)}{\exp\left(\dfrac{2d}{L_a}\right)-2} \times \left\{1 - \frac{2}{\exp\left(\dfrac{2d - \sqrt{\dfrac{2\varepsilon_0\varepsilon_s V_1}{qN_D}}}{L_a}\right)}\right\}$$

$$= \frac{1}{\exp\left(\dfrac{2d}{L_a}\right)-2} \times \left[\exp\left(\dfrac{2d}{L_a}\right)\right.$$

$$\left. -2\exp\left(\frac{2d}{L_a}\right)\exp\left\{\left(-\frac{2d}{L_a}\right) + \frac{\sqrt{\dfrac{2\varepsilon_0\varepsilon_s V_1}{qN_D}}}{L_a}\right\}\right]$$

$$= \frac{1}{\exp\left(\dfrac{2d}{L_\mathrm{a}}\right) - 2} \times \left\{ \exp\left(\frac{2d}{L_\mathrm{a}}\right) - 2\exp\left(\frac{\sqrt{\dfrac{2\varepsilon_0\varepsilon_\mathrm{s}V_1}{qN_\mathrm{D}}}}{L_\mathrm{a}} \right) \right\}$$

$$= \frac{1}{\exp\left(\dfrac{2d}{L_\mathrm{a}}\right) - 2} \times \left\{ \exp\left(\frac{2d}{L_\mathrm{a}}\right) - 2\exp\left(\sqrt{\frac{2\varepsilon_0\varepsilon_\mathrm{s}V_1}{qN_\mathrm{D}L_\mathrm{a}^2}} \right) \right\}$$

$$= \frac{\exp\left(\dfrac{2d}{L_\mathrm{a}}\right) - 2\exp\sqrt{\dfrac{2\varepsilon_0\varepsilon_\mathrm{s}V_1}{qN_\mathrm{D}L_\mathrm{a}^2}}}{\exp\left(\dfrac{2d}{L_\mathrm{a}}\right) - 2} \tag{A6.1.6}$$

APPENDIX 6.2: EQUATION (6.20)

$$\frac{I_\mathrm{L,DR}}{I_\mathrm{L,SS}} = \frac{1 - \dfrac{1}{\cosh(d_\mathrm{N}/L_\mathrm{a})}}{1 - \dfrac{1}{\cosh\{(d_\mathrm{N^-} + d_\mathrm{N})/L_\mathrm{a}\}}} = \frac{\dfrac{\cosh(d_\mathrm{N}/L_\mathrm{a}) - 1}{\cosh(d_\mathrm{N}/L_\mathrm{a})}}{\dfrac{\cosh\{(d_\mathrm{N^-} + d_\mathrm{N})/L_\mathrm{a}\} - 1}{\cosh\{(d_\mathrm{N^-} + d_\mathrm{N})/L_\mathrm{a}\}}}$$

$$= \frac{\cosh(d_\mathrm{N}/L_\mathrm{a}) - 1}{\cosh\{(d_\mathrm{N^-} + d_\mathrm{N})/L_\mathrm{a}\} - 1} \times \frac{\cosh\{(d_\mathrm{N^-} + d_\mathrm{N})/L_\mathrm{a}\}}{\cosh(d_\mathrm{N}/L_\mathrm{a})} \tag{A6.2.1}$$

$$\text{First factor} = \frac{\dfrac{\exp(d_\mathrm{N}/L_\mathrm{a}) + \exp(-d_\mathrm{N}/L_\mathrm{a})}{2} - 1}{\dfrac{\exp\{(d_\mathrm{N^-} + d_\mathrm{N})/L_\mathrm{a}\} + \exp\{-(d_\mathrm{N^-} + d_\mathrm{N})/L_\mathrm{a}\}}{2} - 1}$$

$$= \frac{\exp(d_\mathrm{N}/L_\mathrm{a}) + \exp(-d_\mathrm{N}/L_\mathrm{a}) - 2}{\exp\{(d_\mathrm{N^-} + d_\mathrm{N})/L_\mathrm{a}\} + \exp\{-(d_\mathrm{N^-} + d_\mathrm{N})/L_\mathrm{a}\} - 2}$$

$$= \frac{\exp(d_\mathrm{N}/L_\mathrm{a}) - 2}{\exp\{(d_\mathrm{N^-} + d_\mathrm{N})/L_\mathrm{a}\} - 2} \tag{A6.2.2}$$

ignoring terms with negative powers as explained in Sec.6.3.1 in connection with Eq. (6.12). Similarly,

$$\text{Second factor} = \frac{\dfrac{\exp\{(d_{N^-} + d_N)/L_a\} + \exp\{-(d_{N^-} + d_N)/L_a\}}{2}}{\dfrac{\exp(d_N/L_a) + \exp(-d_N/L_a)}{2}}$$

$$= \frac{\dfrac{\exp\{(d_{N^-} + d_N)/L_a\}}{2}}{\dfrac{\exp(d_N/L_a)}{2}} = \frac{\exp\{(d_{N^-} + d_N)/L_a\}}{\exp(d_N/L_a)} \qquad \text{(A6.2.3)}$$

neglecting negative power terms as before. Applying eqs. (A6.2.3) and (A6.2.2) and re-arranging, Eq. (A6.2.1) becomes

$$\frac{I_{L,DR}}{I_{L,SS}} = \frac{\exp(d_N/L_a) - 2}{\exp(d_N/L_a)} \times \frac{\exp\{(d_{N^-} + d_N)/L_a\}}{\exp\{(d_{N^-} + d_N)/L_a\} - 2}$$

$$= \frac{\exp(d_N/L_a) - 2}{\exp(d_N/L_a)} \times \frac{1}{1 - 2\exp\{-(d_{N^-} + d_N)/L_a\}}$$

$$= \frac{\exp(d_N/L_a) - 2}{\exp(d_N/L_a) - 2\exp(-d_{N^-}/L_a)} \qquad \text{(A6.2.4)}$$

7

DESIGN CONSIDERATIONS OF IGBT UNIT CELL

7.1 SEMICONDUCTOR SELECTION AND VERTICAL STRUCTURE DESIGN

7.1.1 Starting Material

At the genesis of all IGBT design, and in fact the design of any power device, lays the selection of the basic raw material in which the device will be fabricated. The semiconductor material used for the fabrication of present-day IGBTs is silicon, more specifically phosphorus-doped (N-type) Si, either of Czochralski (CZ) or float-zone (FZ) variety. In the CZ process, the polycrystalline silicon contained in a quartz crucible is kept molten by radio-frequency (RF) heating. A seed crystal, suspended in a chuck, is inserted into the melt and withdrawn, resulting in growth of single-crystal silicon at the solid–liquid interface. In the FZ process, a bar of cast silicon is held vertically. By RF heating, a small zone of the bar is melted. The seed crystal is kept at the starting point of the molten zone. Since the single crystal silicon is unsupported, volatile impurities are removed in a reduced pressure ambient. Contaminants introduced from the crucible walls are also absent in FZ silicon. Lower levels of oxygen and carbon contents in the FZ silicon eliminate microprecipitates, the cause of premature breakdown in high-voltage devices.

It will be expedient to examine the reasons underlying the choice of semiconductor material for IGBT fabrication, and explore about any better

---

The Insulated Gate Bipolar Transistor (IGBT) Theory and Design, By Vinod Kumar Khanna.
ISBN 0-471-23845-7 © 2003 Institute of Electrical and Electronics Engineers

Table 7.1 Properties of Power Semiconductor Materials at 300 K

Serial No.	Property	Si	GaAs	4H-SiC
1	Breakdown field (V/cm)	$\sim 3 \times 10^5$	$\sim 4 \times 10^5$	$\sim 3 \times 10^6$
2	Relative permittivity	11.9	12.4	9.7
3	Intrinsic carrier concentration (cm^{-3})	1.08×10^{10}	2.1×10^6	5×10^{-9}
4	Bandgap (eV)	1.12	1.42	3.25
5	Electron mobility (drift) (cm^2/V-sec)	1500	8500	950
6	Electron saturation velocity (cm/sec)	1×10^7	4.4×10^7	2×10^7
7	Minority-carrier lifetime (μsec)	2500	$\sim 10^{-2}$	0.5 to 10^{-3}
8	Thermal conductivity (W/cm-K)	1.5	0.46	4.9

alternative in the future. Material requirements for power IGBTs are as follows: high carrier mobility and high carrier lifetime, for low ON-state voltage drop; high crystal quality, with uniform background dopant concentration; low doping concentrations, to attain high breakdown voltages; high thermal conductivity for rapid heat removal; high melting point and wide bandgap, to allow operation at elevated temperatures. Table 7.1 presents the relevant properties of Si, GaAs, and 4H-SiC (the most attractive polytype) [1,2], three semiconductor materials suitable for IGBT fabrication. SiC exhibits one-dimensional polymorphism. *Polytypes* of SiC, such as 4H-SiC, 6H-SiC, and 3C-SiC, are its one-dimensional polymorphs or polymorphic modifications. *Polymorphs* (allotropes or allotropic modifications) themselves are the various crystalline forms or structures of a substance, element, or compound, possessing different physical properties. Polymorphs of carbon, sulfur and phosphorus are well known.

Apart from its advanced fabrication technology and the ability to produce N-type material by neutron transmutation doping, with phosphorus doping, Si has a moderately wide energy gap and high carrier lifetime. Electron mobility in Si is lower than in GaAs but higher than in SiC. Maximum electric field increases in the order Si, GaAs, SiC, with Si at the lowest extremity.

Having selected the starting material and the type of doping, N-type or P-type (N-type for N-channel IGBT and P-type for P-channel IGBT), the third choice is regarding the preferred orientation of the Si crystal. During semiconductor wafer manufacturing, the surface of the silicon wafer is carefully preoriented to lie along a specific crystallographic plane. This orientation directly affects the device characteristics. For instance, the effective electron mobility in inversion layer in Si [3–5] increases in the order $\mu_e(100) > \mu_e(111) > \mu_e(110)$ because of the anisotropy of conductivity effective mass in Si. The crystal orientation is defined using the *Miller indices* for the plane of atoms. The Miller indices for the crystal plane are obtained by finding the intercepts of the plane on the three rectangular axes in terms of the lattice constant, taking the reciprocal of these numbers, reducing them to the smallest three integers having the same ratio, and enclosing the result in parentheses as (*hkl*). To cite an example, the (100) plane has the intercepts $(1, \infty, \infty)$ on the Cartesian axes. Since the reciprocals of intercepts are $(1, 0, 0)$, the plane is designated by the Miller indices (100). Similarly, the (111) plane has the intercepts $(1, 1, 1)$ while the (110) plane has the intercepts $(1, 1, \infty)$ on the axes.

As an IGBT is turned on by the MOS gate, the ⟨100⟩ orientation is used because the surface state density Q_{ss} for this orientation ($\sim 1\text{–}5 \times 10^{10}$ cm^{-2}) is much lower than for ⟨111⟩ orientation ($10^{11}\text{–}10^{12}$ cm^{-2}), yielding a lower threshold voltage and therefore, smaller driving losses when the ⟨100⟩ orientation is used. To understand the meaning of $Qs_{s,}$ we note that the periodic lattice of the silicon crystal is disrupted at the surface. Consequently, a large number of states are introduced into the forbidden energy gap. These are called *ast surface states*. They are associated with a surface-state charge. They are said to be fast because they can rapidly exchange charges with the substrate semiconductor below.

Theoretically, every silicon atom must contribute one fast state. But since silicon is always covered by oxide—either naturally grown or native oxide, or the oxide grown for device fabrication—the number of these states is typically $10^{10}\text{–}10^{12}$ cm^{-2}, decreasing in the order (111) > (110) > (100) in the ratio 3:2:1. This is one of the reasons why the (100) orientation is preferred for MOSFET and IGBT fabrication. The smallest Q_{ss} in the (100) case gives a low threshold voltage. Furthermore, it has been found that these charges are located within 200 Å from the Si–SiO_2 interface. Their density is a strong function of the oxidation and post-oxidation annealing conditions. Heat treatments in ambient such as H_2 or forming gas at 400–500 °C reduces the surface-state density to 10^{10} cm^{-2}. Hydrogen diffuses to the Si–SiO_2 interfaces and fills the "dangling" missing bonds.

The final fundamental selection to be made is that of the Si doping concentration and thickness. It is common to specify resistivity instead of doping level, as it is easily measurable. The formula for resistivity is

$$\rho = \frac{1}{q(N_D \mu_n + N_A \mu_p)} \tag{7.1}$$

It may be remarked that at a given temperature, the carrier concentration is not necessarily equal to the impurity concentration because all the donor and acceptor impurities may not be ionized.

7.1.2 Breakdown Voltages

The forward and reverse breakdown voltages of an IGBT are limited by avalanche and punchthrough breakdown phenomena and the current amplification by the constituent transistor sections, in the common-base connection (α_{NPN} and α_{PNP}). Transistor gains, α_{NPN} and α_{PNP}, are dependent upon the effective base thickness of the two transistors and the carrier lifetime in the finished device. The design procedure to achieve specified breakdown voltages, leading to a judicious selection of resistivity and thickness, is based on the expressions for breakdown voltages, the avalanche multiplication coefficient, and the base transport factor. As an illustration, to decide the constitution of the N^--base layer for the forward blocking capability, breakdown voltage analysis of the open-base PNP transistor is carried out. Let us first assume that the N^--base thickness is quite large so that the depletion layer formed at the highest voltage to be encountered in operation does not touch the emitter. Then the breakdown of this transistor will be determined only by avalanche phenomenon. The avalanche breakdown voltage is determined by the donor concentration of N^- base and hence the N^--base resistivity. An increase in doping concentration of N^- base results in a decrease in the avalanche breakdown voltage. But in the transistor, the effect of current gain on breakdown voltage cannot be ignored, so overlooking the gain would result in gross oversimplification. Therefore, the effects of finite thickness of N^- base and current gain of the transistor on this voltage are included. Then, punchthrough criterion is determined (see below).

For any diffusion length, open-base transistor breakdown is initiated when the product $M_{\mathrm{p}}\alpha_{\mathrm{T}} = 1$, where M_{p} is the electron multiplication coefficient and α_{T} is the base transport factor. M_{p} varies with reverse bias voltage according to the relation (Appendix 7.1)

$$M_{\mathrm{p}} = \frac{1}{1 - (V/V_{\mathrm{a}})^{\mathrm{n}}} \tag{7.2}$$

where V is the applied reverse voltage, V_{a} is the avalanche breakdown voltage, and $n = 6$ for P^+-N diode and 4 for N^+-P diode. The base transport factor is written as (Appendix 4.8)

$$\alpha_{\mathrm{T}} = \frac{1}{\cosh\left(W/\sqrt{D_{\mathrm{p}}\tau_{\mathrm{p}}}\right)} \tag{7.3}$$

where W is the undepleted base width and D_{p} and τ_{p} are the hole diffusion coefficient and carrier lifetime, respectively. Because the reverse breakdown condition corresponds to $M_{\mathrm{p}} = 1/\alpha_{\mathrm{T}}$, breakdown voltage is given by (Appendix 7.2)

$$V_{\mathrm{BR}} = V_{\mathrm{B}}(1 - \alpha_{\mathrm{T}})^{1/n} \tag{7.4}$$

For abrupt junction approximation, V_B is written as (Appendix 7.3)

$$V_B = 60\left(\frac{N_B}{10^{16}}\right)^{-0.75} \tag{7.5}$$

Abrupt junction is a one-sided or step junction in which impurity concentration on one side is much higher than that on the other side, such as formed by a shallow diffusion or low-energy ion implantation. We take different thicknesses of N^- base, $d_1, d_2, d_3, \ldots$. For each thickness, when the depletion layer width will be close to the P^+ emitter ($\leq$ a diffusion length), injection of holes will occur from emitter toward collector causing punchthrough breakdown. If the diffusion length, L_p, is short, which happens at a low value of carrier lifetime, the depletion layer will stretch over the full N^--base region. Punchthrough breakdown voltage is calculated for $d_1 = 50$

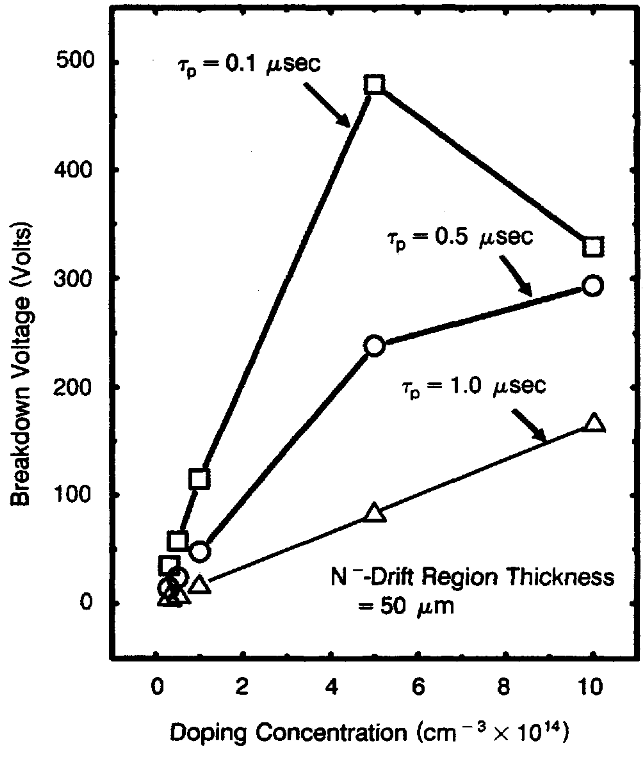


Figure 7.1 Resultant breakdown voltage change with doping concentration. Both avalanche and punchthrough limits are considered. N^--drift region thickness is 50 μm, and hole lifetime (τ_p) in N^- drift region is 0.1, 0.5, and 1 μsec.

μm, $d_2 = 100$ μm and $d_3 = 200$ μm, using the formula (Appendix 7.4)

$$(V_B)_{PT} = \frac{(W_{N^-} - L_p)^2 qN_D}{2\varepsilon_0 \varepsilon_s} \tag{7.6}$$

where $W_{N^-} = d_1, d_2, d_3, \ldots$. This equation is readily obtained from the condition that at punchthrough breakdown, we have [Eq. (A7.4.2)]

$$W_{N^-} = \sqrt{\frac{2\varepsilon_0 \varepsilon_s (V_B)_{PT}}{qN_D}} + L_p \tag{7.7}$$

For any thickness, the breakdown of the IGBT will be determined by both avalanche and punchthrough criteria. Maximum achievable breakdown voltage is represented by the point of intersection of avalanche limit line and punchthrough line for the corresponding thickness. IGBT operation must be

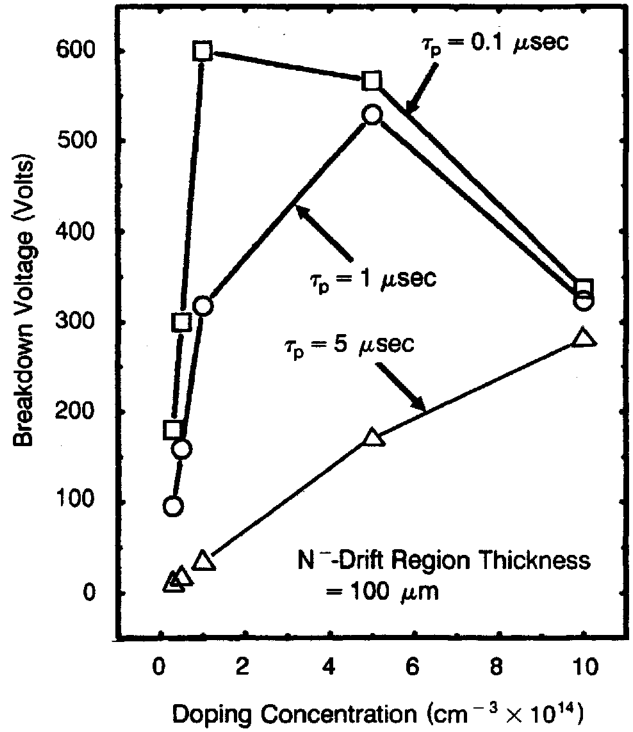


Figure 7.2 Plots showing the effect of doping concentration on the resultant breakdown voltage including both avalanche and punchthrough phenomena. N^--drift region thickness is 100 μm, and hole lifetime (τ_p) in N^--drift region is 0.1, 1, and 5 μsec.

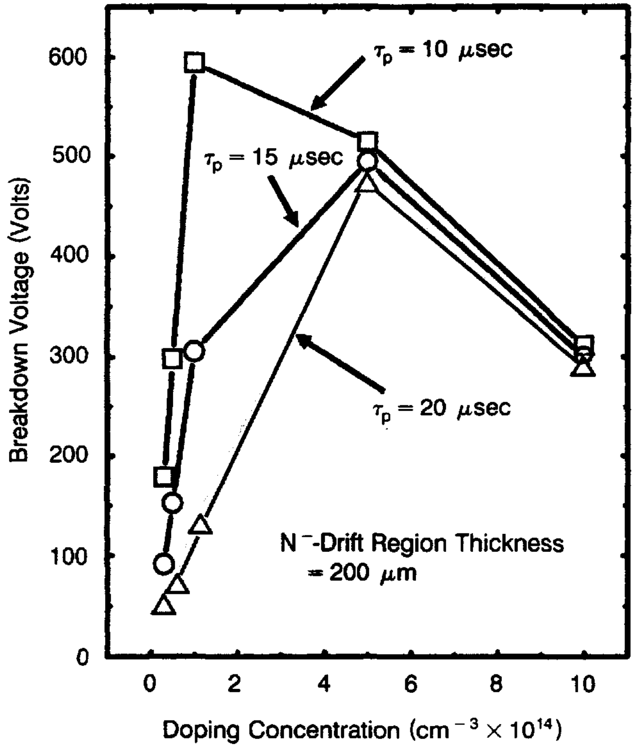


Figure 7.3 Dependence of resultant breakdown voltage on doping concentration. N⁻-drift region thickness is 200 μm, and hole lifetime (τ_p) in N⁻-drift region is 10, 15, and 20 μsec.

confined below the limits set by avalanche and punchthrough mechanisms. Practically, achieved breakdown voltage will critically depend on the carrier lifetime (τ_p) in N^- base governing the diffusion length. This is because $L_p = \sqrt{D_p \tau_p}$, where D_p is the hole diffusion coefficient. The higher the τ_p, the larger the L_p and therefore the lower the punchthrough voltage according to Eq. (7.6).

The variation of resultant breakdown voltage with doping concentration of N^--drift region (and hence resistivity) is shown in Figs. 7.1 to 7.3 for different N^--base thicknesses. In Fig. 7.1, this resultant voltage variation is shown for one W_{N^-} (N^--base thickness) value of 50 μm. It is obtained by taking into account the decrease of undepleted N^--base width with voltage and also combining the dependence of avalanche breakdown voltage (V_a) and punchthrough breakdown voltage $(V_B)_{PT}$ on doping level for different hole lifetime (τ_p) values: 0.1, 0.5, and 1 μsec. These graphs are drawn by selecting the permissible values of breakdown voltage in conformance with both avalanche and punchthrough restrictions. They therefore represent the maximum allowed voltages for various concentrations. Low lifetime values are allowed for this small thickness W_{N^-} of 50 μm, because adequate conductivity modulation can take place, yielding a low forward drop. It is evident from this analysis that doping and thickness of N^- base, along with the post-processing carrier lifetime in the N^- base, decide the breakdown voltage of the IGBT, in terms of both the avalanche and punchthrough considerations. Similar analysis is carried out for $W_{N^-} = 100$ μm in Fig. 7.2. Here higher lifetime values ($\tau_p = 0.1$, 1, and 5 μsec) are used as compared to Fig. 7.1 because at low lifetime values, the conductivity modulation will not be sufficient for a reasonably low ON-state voltage. In Fig. 7.3 the same analysis is repeated for $W_{N^-} = 200$ μm, employing still higher lifetimes ($\tau_p = 10$, 15, and 20 μsec). Keeping in view the aforesaid parameters (namely, resistivity and thickness of N^--base region) and anticipated final carrier lifetime in the IGBT drift region, a reasonable estimate of breakdown voltage resulting from a given N^--base structure can be made. Then by carefully adjustments of these parameters, the N^--base structure of the device can be finalized.

Example 7.1 (a) Choose the N^--base resistivity and thickness of an IGBT for achieving a reverse breakdown voltage of 600 V if the final carrier lifetime in N^- base in the finished device is 2 μsec.

(b) What punchthrough and avalanche breakdown voltages will be achieved if the final carrier lifetime in the N^- base is 10 μsec? Calculate the avalanche breakdown voltage for the current gain at the rated 600-V value. Comment on the results.

Fig. E7.1.1 illustrates the IGBT half-cell for this problem.

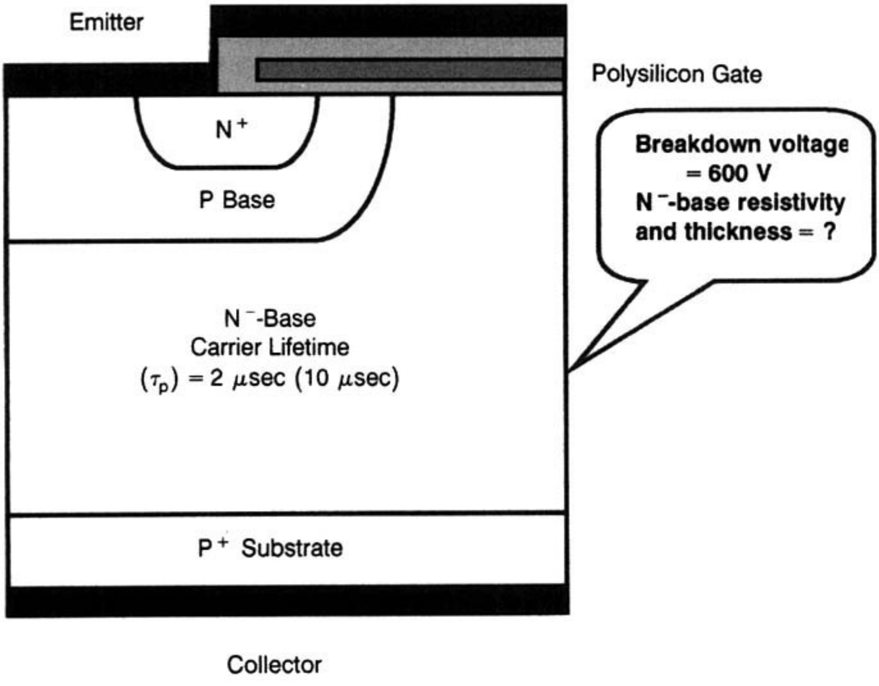


Figure E7.1.1 IGBT structure for Example 7.1.

(a) Since the breakdown voltage will be degraded by current gain α_{PNP}, we choose a higher initial value of avalanche breakdown voltage, 1200 V suppose. For this avalanche breakdown voltage, the doping concentration of N$^-$ base is obtained from the formula (7.5) as $1200 = 60(N_B/10^{16})^{-0.75}$, which gives $N_B = 1.86 \times 10^{14}$ cm^{-3}. For $\tau_p = 2$ μsec, $L_p = \sqrt{12.432 \times 2 \times 10^{-6}} = 49.86$ μm. Keeping the punchthrough breakdown voltage higher than the required value, say 800 V, the thickness of N$^-$ base is calculated from Eq. (7.7) as $d = \sqrt{\{2 \times 8.854 \times 10^{-14} \times 11.9 \times 800/(1.6 \times 10^{-19} \times 1.86 \times 10^{14})\}} + 49.86 \times 10^{-4} = 125.12$ μm. At 800 V, the undepleted base width is 49.86 μm, for which the current gain is $\alpha_{PNP} = 1/\{\cosh(49.86/49.86)\} = 0.64805$ and avalanche breakdown voltage $= 1200\ (1 - 0.64805)^{1/6} = 1007.96$ V. Thus both the current-gain lowered avalanche breakdown voltage and the punchthrough voltage (800 V) are higher than the specified value (600 V). So keeping adequate safety margins, the required N$^-$-base thickness is 125 μm and doping concentration is 1.86×10^{14} cm^{-3}.

(b) For $\tau_p = 10$ μsec, $L_p = 1/(12.432 \times 5 \times 10^{-6}) = 78.84$ μm. From Eq. (7.6), punchthrough voltage $V_{PT} = (125.12 - 78.84)^2 \times 1.6 \times 10^{-19} \times 600/(2 \times 8.854 \times 10^{-14} \times 11.9) = 302.46$ V. At 600 V, the depletion region width $= \{2 \times 8.854 \times 11.9 \times 600/(1.6 \times 10^{-19} \times 1.86 \times 10^{14})\}^{1/2} = 65.18$ μm. Undepleted base width is $125.12 - 65.18 = 59.94$ μm. So, $\alpha_{PNP} = 1/\cosh(59.94/78.84) = 0.767$ and avalanche breakdown voltage decreases to $1200\ (1 - 0.767)^{1/6} = 940.63$ V.

Thus increasing the lifetime from 2 to 5 μsec impairs both punchthrough and avalanche breakdown voltages, but the decrease in punchthrough voltage is more prominent. It is therefore concluded that the designer must have knowledge of the expected final carrier lifetime in N$^-$-base as desired by the turn-off time.

7.1.3 Breakdown Models

Modeling of the avalanche breakdown of semiconductor devices involves the incorporation of a carrier generation model, due to impact ionisation in the group of basic semiconductor equations (Poisson's and electron–hole continuity equations). Generally, the electron and hole current densities J_n, J_p are relatively small in a reverse-biased PN junction. So, analytical expressions are derived for electron and hole concentrations, as functions of electrostatic potential. These expressions are substituted into Poisson's equation, and the resulting nonlinear Poisson's equation is solved to determine the potential throughout the device. This model is further simplified by using the *depletion approximation.* In the depletion approximation, it is postulated that the space-charge region consists entirely of the charges of ionized acceptors and donors, so that the electron and hole concentrations n and p are insignificantly smaller than the ionized dopant concentrations. Also, quasi-neutrality conditions prevail in the undepleted regions of the device. In symbolic form, in a PN diode with depletion region boundary points $-x_P$ and $+x_N$: (i) For $0 \geq x \geq -x_P$, the charge density $\rho = -qN_A^-$, because acceptor concentration N_A is much greater than electron concentration on P side (n_P) or hole concentration on P side (p_P). (ii) For $x_N \geq x \geq 0$, the charge density $\rho = +qN_D^+$, because donor concentration N_D is much greater than electron concentration on N side (n_N) or hole concentration on N side (n_P). (iii) For

$x > x_N$ and $x < -x_P$, charge density ρ is zero. This means that in the bulk P and N regions, (i.e., regions outside the depletion region on P and N sides), $\rho = 0$.

Mesh specification files are generally divided into two sections: mesh skeleton and mesh refinement criteria. *Mesh skeleton* contains the hierarchical declaration of points, edges, and regions. Two coordinates enclosed within parentheses define a point. Points define edges, whereas regions are defined by edges. *Mesh refinement criteria* specify the maximum and minimum spacing between nodes, the number of refinement divisions, and the conditions of refinement.

The electric field calculations are carried out by iteratively solving a set of difference equations representing an approximation to the Poisson's equation, along with an algorithm that finds which regions of the device have space charge. In the first step, the Poisson's difference equations are solved assuming a charge density equal to the net charge of ionized impurities. But during reverse-bias operation, the regions of the device near the contacts are neutral. The electrostatic potentials of the regions around anode and cathode of the PIN diode are fixed by the anode and cathode potentials respectively. Depending on the side of the junction, the potential of these regions is either zero or equal to the applied voltage. In the second step, the value of the computed electrostatic potential at a mesh point is restricted from becoming less than zero (anode potential) or more than the applied voltage (cathode potential). It is clamped to the appropriate contact potential, depending upon whether it is greater than the cathode potential or less than the anode potential. This clamping operation is referred to as *depletion region logic*. The logic accounts for the interaction of the mobile carriers with the potential distribution and determines the regions of the device having space charge. Clamping fulfills the objective of obtaining solutions to Poisson's equation using the charge density in the depletion region and constant potential solutions of Laplace's equation outside this region. Effectively, the procedure is that Poisson's equation is solved in the depletion region with charge density given by impurity concentration and Laplace's equation is solved in the quasi-neutral bulk region. Once the solution has been obtained, the depletion region boundaries are determined by locating the lines separating the regions of the device at constant potential and those at varying potential. It may be noted that the restriction imposed on the solution to lie between zero and the applied voltage is valid because local or higher voltages will produce regions of local potential minima for holes or electrons, respectively. Then the free carriers, which are present everywhere in the device, will travel to these local minima. This will continue until the potential has readjusted itself to the applied voltage.

The above algorithm is successively applied at each mesh point of the domain of simulation. This operation is termed "sweeping." As long as each node (mesh point in the simulation domain) is updated during the sweep, the order of handling of the nodes is inconsequential. After each sweep, the maximum change in potential is compared to a predetermined accuracy limit; and if the maximum change exceeds this limit, the sweeping action is repeated. In this manner, the electrostatic potential is determined with required accuracy.

The method of finding the breakdown voltage follows an indirect approach. It involves the evaluation of the electron and hole multiplication coefficients M_e and M_h that determine the number of secondary electron–hole pairs produced by impact ionization by one electron or hole when it traverses from one end of the depletion region to the other [6, 7]. At breakdown, either $M_e \to 0$ or $M_h \to 0$. The coefficients approach infinity when the integrals in their denominators progress toward unity. These integrals known as the electron and hole ionization integrals I_e and I_h are given by

$$I_e = \int_{x_n}^{x_p} \alpha_n(x) \exp\left[\int_{x_n}^{x_p} \{\alpha_p(x') - a_n(x')\}\, dx'\right] dx \tag{7.8}$$

$$I_h = \int_{x_n}^{x_p} \alpha_p(x) \exp\left[\int_{x_a}^{x_p} \{\alpha_n(x') - a_p(x')\}\, dx'\right] dx \tag{7.9}$$

where α_n, α_p are the electron and hole ionization coefficients, defined as the number of electron–hole pairs produced by an electron or hole traversing a unit distance through the depletion region along the direction of the electric field.

Although the ionization integrals can be evaluated along any path through the depletion region, the critical ionization path is the trajectory of the electric field line that gives the largest ionization integral value. Practically, the breakdown voltage is measured by the electron or hole ionization integral becoming unity when determined along the critical ionization path. Furthermore, it is not necessary to evaluate both these integrals. For a P^+N junction, $I_h > I_e$ and the reverse is true for an N^+P junction. The accuracy of breakdown voltage calculation depends upon that of the ionization integral calculation, and hence on the ionization integral models.

Example 7.2 For a parallel-plane abrupt N^+P^- junction, solve Poisson's equation $dE/dx = -qN_A/\varepsilon_0\varepsilon_s$, where E is the electric field, x is the distance, ε_0 is free-space permittivity, ε_s is dielectric constant of silicon, and N_A is the doping concentration on the P^- side. Using the formula $\alpha = 1.8 \times 10^{-35}E^7$ for the dependence of the impact ionization coefficient on the field E, apply the condition $\int_0^{W_c} \alpha\, dx = 1$ (W_c = width of the depletion region at breakdown) to determine W_c. Hence derive the equation for the avalanche breakdown voltage of an abrupt parallel-plane junction.

Integration of Poisson's equation yields

$$\int \frac{dE}{dx} = -\int \frac{qN_A}{\varepsilon_0\varepsilon_s} \quad \text{or} \quad \int dE = -\frac{qN_A}{\varepsilon_0\varepsilon_s}\int dx \quad \text{or} \quad E = -\frac{qN_A x}{\varepsilon_0\varepsilon_s} + \text{Constant} \tag{E7.2.1}$$

From the boundary condition $E = 0$ at $x = W$, we have

$$E = -\frac{qN_A W}{\varepsilon_0 \varepsilon_s} + \text{Constant} \quad \text{or} \quad \text{Constant} = \frac{qN_A W}{\varepsilon_0 \varepsilon_s} \tag{E7.2.2}$$

Putting the value of the constant from Eq. (E7.2.2) into Eq. (E7.2.1), we get

$$E(x) = \frac{qN_A}{\varepsilon_0 \varepsilon_s}(W - x) \tag{E7.2.3}$$

Since $E = -dV/dx$ and qN_A is negative, Eq. (E7.2.3) is written in terms of potential as

$$\frac{dV}{dx} = \frac{qN_A}{\varepsilon_0 \varepsilon_s}(W - x) \tag{E7.2.4}$$

Integrating Eq. (E7.2.4), we get

$$\int dV = \frac{qN_A}{\varepsilon_0 \varepsilon_s}\left\{W\int dx - \int x\,dx\right\} + \text{Constant} \quad \text{or} \quad V = \frac{qN_A}{\varepsilon_0 \varepsilon_s}\left(Wx - \frac{x^2}{2}\right) + \text{Constant} \tag{E7.2.5}$$

Since the potential is zero in the N$^+$ region (i.e., $V = 0$ at $x = 0$), Constant = 0 and Eq. (E7.2.5) reduces to

$$V = \frac{qN_A}{\varepsilon_0 \varepsilon_s}\left(Wx - \frac{x^2}{2}\right) \tag{E7.2.6}$$

Now, the ionization coefficient α is related to the electric field E as

$$\alpha = 1.8 \times 10^{-35} E^7 \tag{E7.2.7}$$

Putting the expression for E from Eq. (E7.2.3) into Eq. (E7.2.7), the ionization integral equals the unity condition for breakdown, that is; $\int \alpha\, dx = 1$ is

$$\int_0^{W_c} \left[1.8 \times 10^{-35}\left\{\frac{qN_A}{\varepsilon_0 \varepsilon_s}(W - x)\right\}^7\right] dx = 1 \tag{E7.2.8}$$

where W_c is the critical depletion region width at breakdown. Eq. (E7.2.8) may be rewritten as

$$-1.8 \times 10^{-35} \frac{q^7 N_A{}^7}{(\varepsilon_0 \varepsilon_s)^7} \int_0^{W_c} (W - x)^7 dx = 1 \tag{E7.2.9}$$

or

$$\int_0^{W_c}(W-x)^7\,dx = -\frac{1\times(8.854\times10^{-14}\times11.9)^7}{1.8\times10^{-35}\times(1.6\times10^{-19})^7}\times\frac{1}{N_A{}^7}$$

$$= -2.9833\times10^{82}\times\frac{1}{N_A{}^7} \tag{E7.2.10}$$

because of the negative sign of qN_A. Expanding the expression $(W-x)^7$ by the binomial theorem, we obtain

$$\int_0^{W_c}\left\{W^7-7W^6x+\frac{7(7-1)}{2!}W^5x^2-\frac{7(7-1)(7-2)}{3!}W^4x^3+\cdots-x^7\right\}dx$$

$$= -\frac{2.9833\times10^{82}}{N_A{}^7} \tag{E7.2.11}$$

Neglecting lower powers of x, we obtain

$$\int_0^{W_c}x^7\,dx = \frac{2.9833\times10^{82}}{N_A{}^7} \tag{E7.2.12}$$

$$\text{or,}\quad\left[\frac{x^8}{8}\right]_0^{W_c} = \frac{2.9833\times10^{82}}{N_A{}^7}\quad\text{or}\quad\frac{W_c^8}{8}-0=\frac{2.9833\times10^{82}}{N_A{}^7} \tag{E7.2.13}$$

$$\therefore W_c = (2.9833\times10^{82}\times8)^{1/8}\times N_A^{-7/8} = 2.644\times10^{10}N_A{}^{-7/8} \tag{E7.2.14}$$

But from Eq. (E7.2.6), at breakdown we have $x = W_c$; hence

$$V=\frac{qN_A}{\varepsilon_0\varepsilon_s}\left(W_c^2-\frac{W_c^2}{2}\right)=\frac{qN_AW_c{}^2}{2\varepsilon_0\varepsilon_s} \tag{E7.2.15}$$

Substituting for W_c from Eq. (E7.2.14) into Eq. (E7.2.15), we have

$$V=\frac{qN_A(2.644\times10^{10}N_A^{-7/8})^2}{2\varepsilon_0\varepsilon_s}=\frac{1.6\times10^{-19}\times(2.644\times10^{10})^2}{2\times8.854\times10^{-14}\times11.9}N_A{}^{1-7/4}$$

$$=5.31\times10^{13}N_A{}^{-3/4} \tag{E7.2.16}$$

Hence, breakdown voltage of the parallel-plane junction is

$$BV_{pp}=5.31\times1o^{13}N_A{}^{-3/4} \tag{E7.2.17}$$

7.2 IGBT DESIGN BY ANALYTICAL CALCULATIONS AND NUMERICAL SIMULATIONS

7.2.1 Design Approach and CAD Simulation Hierarchy

Like breakdown voltage, each parameter of the IGBT needs attention. Extensive analytical calculations and simulation studies are performed before freezing the design. Electrical parameters are calculated using analytical equations as well as the 2-D/3-D simulator.

The data sheet features the voltage and current ratings and switching speeds. Main specifications of the IGBT are (a) the *absolute maximum ratings* such as collector–emitter voltage, continuous and pulsed collector–emitter current, power dissipation, operating and storage junction temperature range, and saturation voltage, (b) *electrical characteristics* like collector–emitter breakdown voltage, gate–threshold voltage, forward and reverse gate–body leakage currents, zero-gate voltage collector current, and forward transconductance, (c) *capacitances* such as input, output, and reverse transfer capacitance, (d) *switching times* such as, turn-on delay time, rise time, turn-off delay time, and fall time, and (e) *thermal characteristics* like junction-to-case, case-to-sink, and junction-to-ambient thermal resistances.

The design of an IGBT involves the optimization of various physical and process parameters for achieving the required electrical characteristics and specifications. Given a set of specifications, device design aims at relating (a) the terminal electrical characteristics of an IGBT with its internal structure and (b) the physical mechanisms involved in its functioning. IGBT design includes: design of physical and electrical parameters, for the extraction of parameters for geometric layout; process simulation for the doping profile, oxide, channel length, polySi gate thickness, and so on; layout generation using CAD methods, and IGBT design verification. Process simulation involves (a) optimization of ion implantation dose and energy for N- and P-type impurity junction depths and surface concentrations, (b) characterization of gate and field oxide for thickness, (c) quality in terms of surface-state density Q_{SS}, and (d) polySi deposition temperature and time to obtain required thickness. Only device design will be dealt with here, while process design will be looked into in Chapter 9.

Like very large scale integrated circuits, the design of power semiconductor devices such as DMOSFET, IGBT, and MCT, along with associated circuitry, involves advanced CAD tools employing scientific workstations. So, power device design resembles VLSI design. Furthermore, simulation plays a vital role in IGBT design enabling the designer to study the viability of the designed chip without resorting to costly manufacturing processes. Also, it is possible to investigate the influence of environmental parameters, like temperature, on the chip before fabrication. Thus simulation allows the checking of functionality and performance of the chip as well as the ambient effects on it. Necessary corrective measures are then applied and the design is modified.

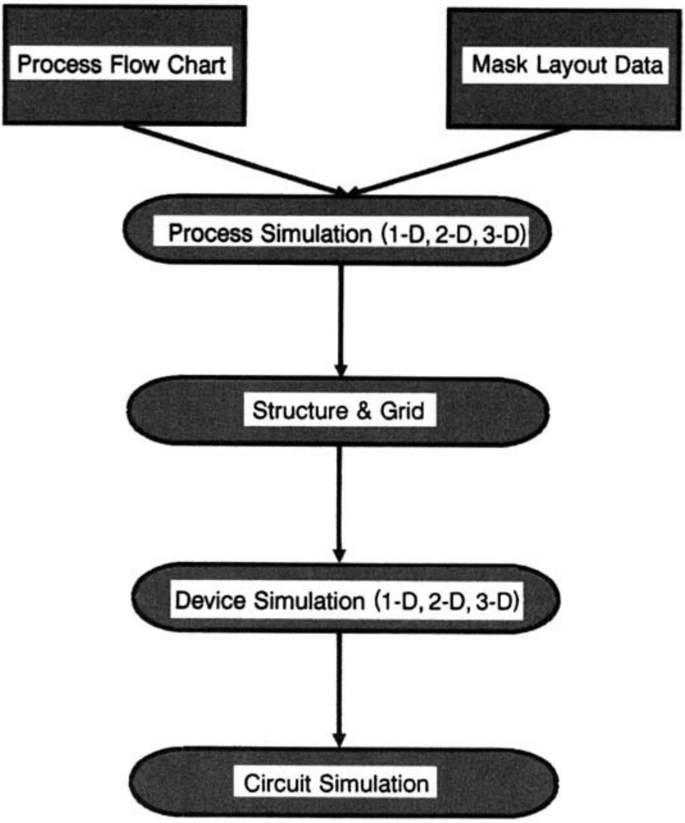


Figure 7.4 Rungs of the simulation ladder.

The simulation and design hierarchy for the IGBT and all power devices is shown in Fig. 7.4. The starting inputs are the process flow chart and mask layout data, which are fed to the process simulator. Other key levels are device and circuit simulation. For circuit simulation, the list of components (netlist) is entered. These components are connected to one or more nodes. The simulator for carrying out the computations uses models describing the operation of these components, mainly resistive elements (e.g., resistors, current, and voltage sources) and energy storage elements (e.g., capacitors and inductors). Device simulation relies on the quality of the physical models formulated for the semiconductor devices. In both device and process simulation, differential equations are solved numerically in uni- or multiple dimensions. The purpose of these simulations is to forecast the electrical characteristics of the device and the required optimized processing.

After the process simulation has been done, the relevant geometrical data and doping profiles are transferred to the device simulator. The structure of the device equations makes it necessary that physically based finite-element discretisations, conforming to certain restrictions, be applied. Hence regridded results from process simulation are used. Device simulation software such as PISCES [8, 9], BAMBI [10], MEDICI [11], DESSIS [12], and so on, are widely used in semiconductor industry. Steady-state and transient characteristics are simulated; and with circuit simulators, gate drive circuits, snubbers, and protection circuits are exhaustively simulated.

A primary advantage of DMOSFET technology used in IGBTs, is the precise control of channel length and the realization of small channel lengths irrespective of photolithographic constraints. Small channel length gives high transconductance (g_m), low channel resistance, and high aspect ratio [channel width (W)/channel length (L)], and therefore high collector current, for a given channel width, as per device design. The collector current also depends on the aspect ratio (W/L) and α_{PNP}. The current gain α_{PNP} should be as high as possible, but traded off with breakdown voltage and latching current density. For fixed α_{PNP} and channel length, the channel width is varied to get the required current. To increase the current capability and to reduce the ON resistance, the cell array concept is applied in which thousands of cells are connected in parallel. All cells are connected in parallel, thus reducing the ON resistance. The number of cells depends on the total current and size of one cell. Cell structure is optimized to increase the cell packing density, and thereby the current capability. To reduce the drive energy of the IGBT, its transconductance g_m has to be increased, which requires small channel length and high α_{PNP}.

To begin, different epilayer doping and thickness along with different field ring spacings are used to optimize the drift region of the device for achieving the required breakdown voltage. When the field ring structure is employed, the reverse voltage is shared among the main junction and the guard rings. After optimizing the drift region, the channel region is evaluated for different P-base doping level and thicknesses. Optimization of the P base is critical because it decides the threshold voltage V_{Th}, transconductance g_m, input capacitance, and so on. Doping level of the P base is chosen higher than the drift region, to avoid punchthrough. Surface concentration in the P base and gate oxide thickness are the main deciding factors, for threshold voltage. Threshold voltage is optimized for different values of P-base surface concentration and gate oxide thickness. Other device parameters such as channel conductance, ON-resistance, power dissipation, frequency response, and so on, are then calculated. Finally, the electron irradiation process is evaluated and optimized with respect to beam energies and dose, to control the carrier lifetime.

7.2.2 Design Software

Several software packages are available to arrive at a realistic and satisfactory design. Integrated Systems Engineering Technology Computer Aided Design (ISE-TCAD) tools is one such package [13]. This software comprises several modules, such as MESH-ISE for grid generation, DESSIS-ISE, a 1-D, 2-D, and 3-D device and circuit simulator, DIOS-ISE for process simulation, and visualization tools like INSPECT and PICASSO.

In the simulation package ISE-TCAD, the basic semiconductor equations (Poisson's and electron–hole continuity equations) are numerically solved on the mesh points mapped over the device structure under study. The numerical solution converges in accordance with the defined criteria for analysis. The typical input consists of structure of the device, mesh size, impurity profiles in different regions, and physical effects and models for performing the simulation. The input is given in the form of files, namely, the msh.bnd or *mesh boundary file* for defining the geometry of the device under examination; msh.cmd or *mesh command file* for impurity profiles in the structure and mesh refinements to focus at the critical regions of the device; or des.cmd or *dessis command file* for specifying the voltages/currents applied on the electrodes of the device, physical models to be used in the simulations, the permissible errors in the parameters, the number of iterations to be done, and the method of solving Poisson's equation (e.g., *coupled* employing full Newton method and *plug-in* with Gummel iteration), along with the approach to problem solving, whether quasi-stationary, transient, and so on. "Quasi-stationary command" is used to ramp a device from one solution to another by modifying its boundary conditions (e.g., ramping of the voltage at a contact). "Transient command" is used to run a transient solution.

7.2.3 Physical Models in DESSIS-ISE

In ISE package, the module for device and circuit simulations is *DESSIS-ISE*. In this module, three simulation modes are offered: (i) *Drift-diffusion model* for isothermal simulations with stationary transport in low power density devices with long active regions. "Stationary transport" means that only the carriers move, not any conducting media. (ii) *Thermodynamic model* taking into account self-heating effects in high-power-density devices with long active regions. (iii) *Hydrodynamic model* to account for self-heating and nonstationary transport, suited to devices with small active regions. "Nonstationary transport" implies that physical movement of the medium is considered along with carrier motion.

In the drift-diffusion model, the three governing equations for charge transport are [12]:

Poisson's Equation:
$$\nabla \cdot \varepsilon_s \nabla \psi = -q(p - n + N_{D^+} - N_{A^-}) \tag{7.10}$$

Electron Continuity Equation:
$$\nabla \cdot \vec{J}_n = qR + q\frac{\partial n}{\partial t} \tag{7.11}$$

Hole Continuity Equation:
$$\nabla \cdot \vec{J}_p = qR + q\frac{\partial p}{\partial t} \tag{7.12}$$

where ε_s is the relative permittivity of Si, ψ is the electric potential, q is the electronic charge, n and p are the electron and hole concentrations, N_{D^+} and N_{A^-} are the ionized donor and acceptor concentrations, J_n and J_p are the electron and hole current densities, and R is the net electron–hole recombination rate. The thermodynamic model is an extension of the drift-diffusion approach, taking into consideration the electrothermal effects assuming that the charge carriers are in thermal equilibrium with the lattice, and the hydrodynamic model takes a step further with the inclusion of nonstationary effects.

The dependence of effective intrinsic carrier concentration, n_{ie}, on net impurity concentration (N) and temperature (T) is described by the equation [14]

$$n_{ie}^2(N,T) = n_{i0}^2(T)\exp\left\{\frac{q\Delta V_{g0}(N)}{kT}\right\} \tag{7.13}$$

where $n_{i0}(T)$ is the intrinsic carrier concentration for low concentrations, varying only with temperature:

$$n_{i0}^2(T) = C_1 T^3 \exp\left(-\frac{qV_{g0}}{kT}\right) \tag{7.14}$$

with the constants $C_1 = 9.61 \times 10^{32}$ and $V_{g0} = 1.206$ V. The experimentally derived bandgap narrowing $\Delta V_{g0}(N)$ caused by degenerate doping is fitted by the following empirical formula

$$\Delta V_{g0}(N) = V_1\left(F + \sqrt{F^2 + C_2}\right) \tag{7.15}$$

where the constants are $V_1 = 9 \times 10^{-3}$ V, $F = \ln(N/N_0)$, $N_0 = 10^{17}$ cm^{-3}, and $C_2 = 0.5$.

Bandgap variation with temperature is given by [13]

$$E_g(T) = E_g(0) - \frac{\alpha T^2}{T + \beta} \tag{7.16}$$

where $\alpha = 4.73 \times 10^{-4}$ eV/K, $\beta = 636$ K, and $E_g(0) = 1.15$–1.16 eV.

Mobility is modeled in a modular way. The bulk mobility is a temperature-dependent constant given by

$$\mu_{\text{const}} = \mu_L\left(\frac{T}{300}\right)^{-\zeta} \tag{7.17}$$

For electrons, $\mu_L = 1417$ cm^2/V-sec and $\zeta = 2.5$; for holes, $\mu_L = 470.5$ cm^2/V-sec and $\zeta = 2.2$. The subscript L pertains to the lattice so that the model applies to bulk mobility in undoped materials only. The model of Masetti et al. describes impurity scattering, extending the Caughey–Thomas equation to the heavy doping range [15–18],

$$\mu_{\text{dop}} = \mu_{\text{min1}} \exp(-P_c/n_i) + \frac{\mu_{\text{const}} - \mu_{\text{min2}}}{1 + (n_i/C_r)^{\alpha}} - \frac{\mu_1}{1 + (C_s/n_i)^{\beta}} \tag{7.18}$$

where $\mu_{\text{min1}} = 68.5$ cm^2/V-sec for P and 44.9 cm^2/V-sec for B, $\mu_{\text{min2}} = 68.5$ cm^2/V-sec for P and 0 for B, $\mu_1 = 56.1$ cm^2/V-sec for P and 29.0 cm^2/V-sec for B, $C_r = 9.2 \times 10^{16}$ cm^{-3} for P and 2.23×10^{17} cm^{-3} for B, $C_s = 3.41 \times 10^{20}$ cm^{-3} for P and 6.10×10^{20} cm^{-3} for B, $\alpha = 0.711$ for P and 0.719 for B, $\beta = 1.98$ for P and 2.0 for B, and $P_c = 0$ and 9.23×10^{16} cm^{-3} for P and B, respectively.

To include the degradation of mobility at interfaces such as in IGBT channels, the bulk mobility (μ_b) is combined with mobilities, surface acphonon scattering (μ_{ac}), and surface roughness scattering (μ_{sr}) by the Matthiessen rule defined by

$$\frac{1}{\mu} = \frac{1}{\mu_b} + \frac{1}{\mu_{ac}} + \frac{1}{\mu_{sr}} \tag{7.19}$$

Mobility with carrier–carrier or electron–hole scattering (μ_{eh}) can also be added by the same rule.

Recombination through deep impurity levels, known as the Shockley–Read–Hall (SRH) recombination, is expressed by the well-known equation [19]

$$R_{\text{net}}^{\text{SRH}} = \frac{np - n_{\text{i,eff}}^{2}}{\tau_p(n + n_1) + \tau_n(p + p_1)} \tag{7.20}$$

where $n_1 = n_{\text{i,eff}} \exp\{E_{\text{trap}}/(kT)\}$, $p_1 = n_{\text{i,eff}} \exp\{-E_{\text{trap}}/(kT)\}$, E_{trap} is the energy difference between the defect level and intrinsic level, and minority-carrier lifetimes τ_n and τ_p are modeled as the products of doping-dependent, field-dependent and temperature-dependent factors.

For breakdown analysis, Poisson's equation is solved and ionization integrals are obtained. Inspection of the ionization integrals is used to identify junction breakdown by avalanche generation. The ionization integrals are computed along the electric field lines through the depletion zone. The output gives the information for the ionization integrals for all computed paths with increasing voltage, applied in steps. Avalanche breakdown is deemed to occur when an ionization integral equals unity. The quasi-stationary simulation is therefore terminated as soon as the largest ionization integral is unity. The values of electron and hole ionization integrals as well as the mean ionization integral are stored for visualization. A predetermined algorithm makes the initial guess for electrostatic potential ψ, and the quasi-Fermi potentials for electrons and holes.

For simulations, the device is defined in the DESSIS input file, by its mesh and doping in the *file section*; its contacts are defined in the *electrode and interface section*; the selected physical models are defined in the *physics section*; and the results to be saved are defined in the *plot section.*

7.2.4 Calculation and Simulation Procedure

A power IGBT essentially comprises a repetitive array of cells arranged in a topological layout, providing a large channel aspect ratio. Therefore, to understand the electrical behaviour of the devices, it is necessary to simulate the performance of the unit cell in terms of physical parameters—for example, geometry, impurity profiles, and so on [20, 21]. For illustrative purposes, representative IGBT cells for medium- and high-voltage operation have been used with typical dimensional and profile data based on forward current, blocking capability, and technological considerations. These data have been finalized from first-order analytical calculations.

Emitter and Base Doping Profiles, Channel Length, Forward Voltage Drop and Transconductance. For the 2.4-μm-channel-length, high-voltage IGBT cell (Fig. 7.5), the P-base depth is 4 μm and the N^+ source depth is 1 μm, yielding a channel length of 2.4 μm. The N^- epilayer concentration has been taken as 6.61×10^{13} cm^{-3} ($\rho = 70$ Ω-cm) and thickness is 150 μm. A 2.4-μm-channel-length, medium-voltage IGBT cell has an upper structure identical to that of the high-voltage cell. The only difference between the two

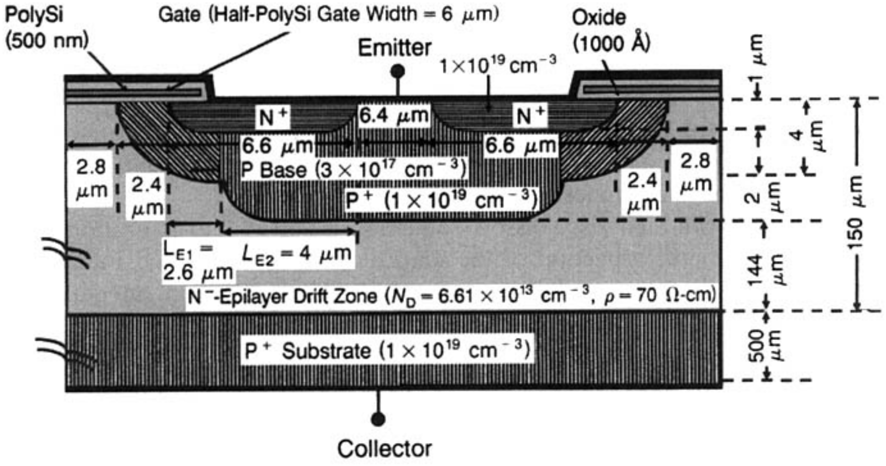

PolySi
(500 nm)
Gate (Half-PolySi Gate Width = 6 μm)
Emitter
1×10¹⁹ cm⁻³
Oxide
(1000 Å)
N⁺
6.4 μm
N⁺
6.6 μm
6.6 μm
P Base (3 × 10¹⁷ cm⁻³)
P⁺ (1 × 10¹⁹ cm⁻³)
2.8 μm
2.4 μm
2.4 μm
2.8 μm
L_E1 = 2.6 μm
L_E2 = 4 μm
N⁻-Epilayer Drift Zone (N_D = 6.61 × 10¹³ cm⁻³, ρ = 70 Ω-cm)
P⁺ Substrate (1 × 10¹⁹ cm⁻³)
1 μm
4 μm
2 μm
144 μm
150 μm
500 μm
Collector

Figure 7.5 Cross-sectional representation of a 2.4-μm-channel-length high-voltage IGBT cell.

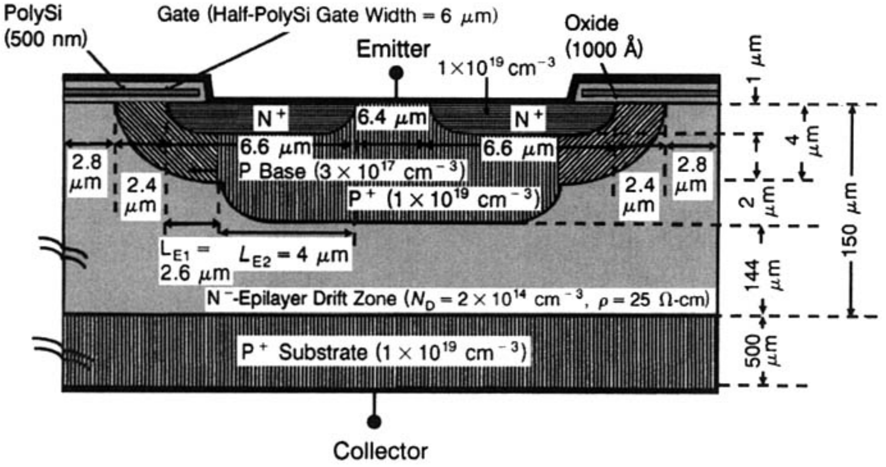

PolySi
(500 nm)
Gate (Half-PolySi Gate Width = 6 μm)
Emitter
1×10¹⁹ cm⁻³
Oxide
(1000 Å)
N⁺
6.4 μm
N⁺
6.6 μm
6.6 μm
P Base (3 × 10¹⁷ cm⁻³)
P⁺ (1 × 10¹⁹ cm⁻³)
2.8 μm
2.4 μm
2.8 μm
2.4 μm
L_E1 = 2.6 μm
L_E2 = 4 μm
N⁻-Epilayer Drift Zone (N_D = 2 × 10¹⁴ cm⁻³, ρ = 25 Ω-cm)
P⁺ Substrate (1 × 10¹⁹ cm⁻³)
1 μm
4 μm
2 μm
144 μm
150 μm
500 μm
Collector

Figure 7.6 Cross-sectional representation of a 2.4-μm-channel-length medium-voltage IGBT cell.

cells lies in the resistivity and starting thickness of epilayer. For the 2.4-μm-channel length, medium-voltage IGBT cell (Fig. 7.6), starting epilayer thickness was 60 μm and doping was 2×10^{14} cm^{-3} ($\rho = 25$ Ω-cm).

An IGBT cell with a channel length of 0.96 μm has been simulated to obtain a low forward voltage drop. Achieving a low initial forward drop is essential considering that the voltage drop will increase when the IGBT is subjected to irradiation to decrease its turn-off time. In the 0.96-μm-channel-length, high-voltage IGBT cell (Fig. 7.7), the N$^+$ region depth is

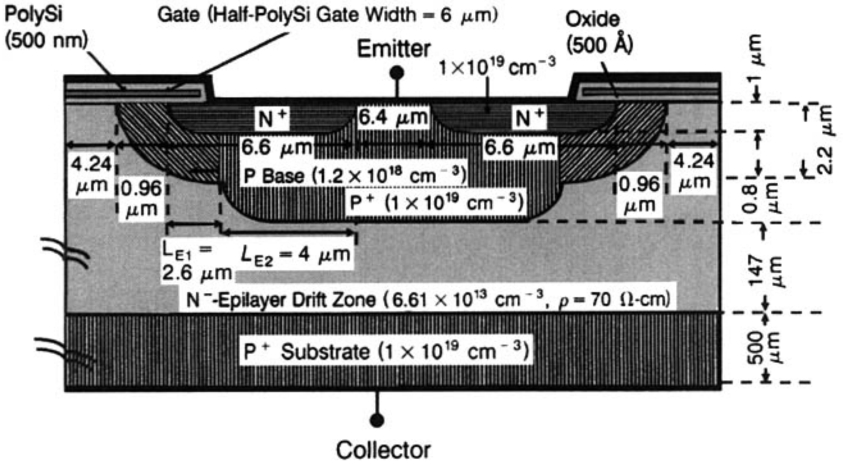

PolySi
(500 nm)
Gate (Half-PolySi Gate Width = 6 μm)
Emitter
1×10¹⁹ cm⁻³
Oxide
(500 Å)
N⁺
6.4 μm
N⁺
6.6 μm
6.6 μm
P Base (1.2 × 10¹⁸ cm⁻³)
P⁺ (1 × 10¹⁹ cm⁻³)
4.24 μm
0.96 μm
4.24 μm
0.96 μm
L_E1 = 2.6 μm
L_E2 = 4 μm
N⁻-Epilayer Drift Zone (6.61 × 10¹³ cm⁻³, ρ = 70 Ω-cm)
P⁺ Substrate (1 × 10¹⁹ cm⁻³)
1 μm
2.2 μm
0.8 μm
147 μm
500 μm
Collector

Figure 7.7 Two-dimensional diagram of a 0.96-μm-channel-length high-voltage IGBT cell.

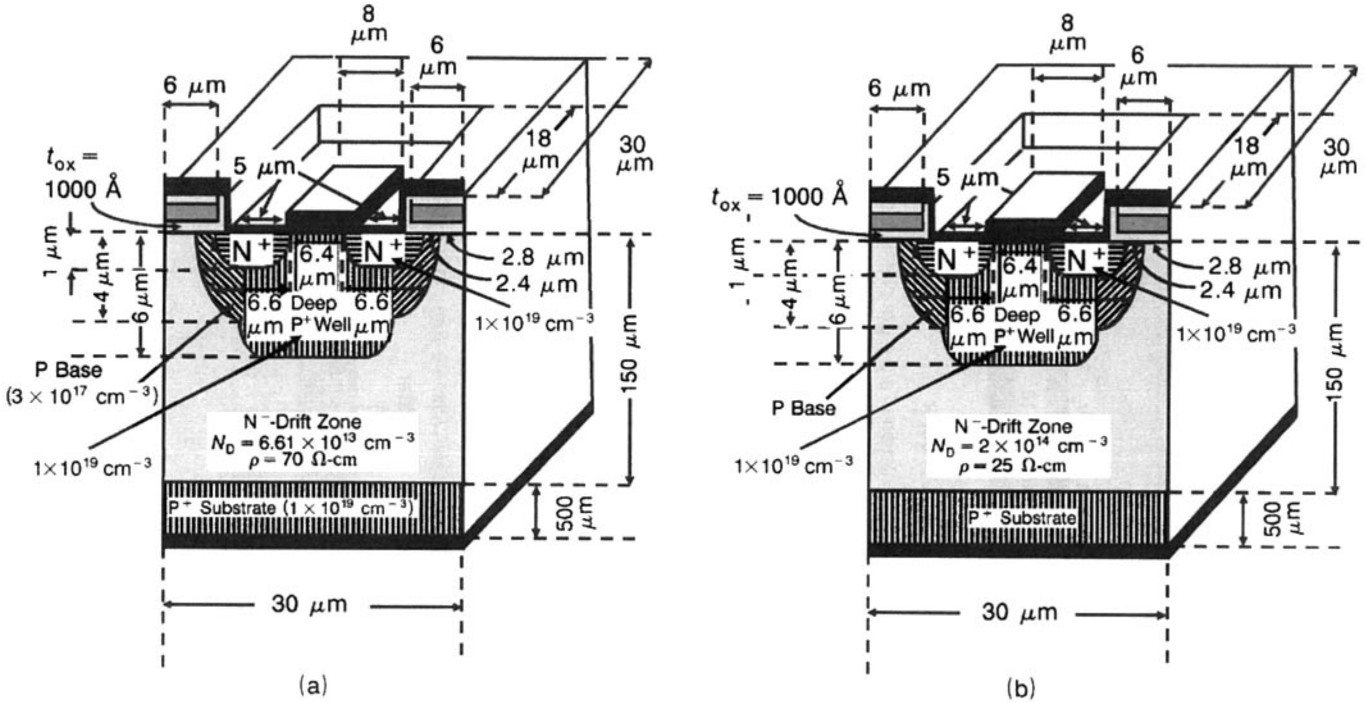

8 μm
6 μm
6 μm
t_{ox} = 1000 Å
5 μm
18 μm
30 μm
1 μm
4 μm
6 μm
N⁺
N⁺
6.4 μm
6.6 μm
Deep P⁺ Well
6.6 μm
2.8 μm
2.4 μm
1×10¹⁹ cm⁻³
P Base
(3 × 10¹⁷ cm⁻³)
1×10¹⁹ cm⁻³
N⁻-Drift Zone
N_D = 6.61 × 10¹³ cm⁻³
ρ = 70 Ω-cm
P⁺ Substrate (1 × 10¹⁹ cm⁻³)
150 μm
500 μm
30 μm
(a)
N_D = 2 × 10¹⁴ cm⁻³
ρ = 25 Ω-cm
P⁺ Substrate
(b)

Figure 7.8 Three-dimensional view of 2.4-μm-channel-length IGBT cells: (a) High-voltage cell and (b) medium-voltage cell.

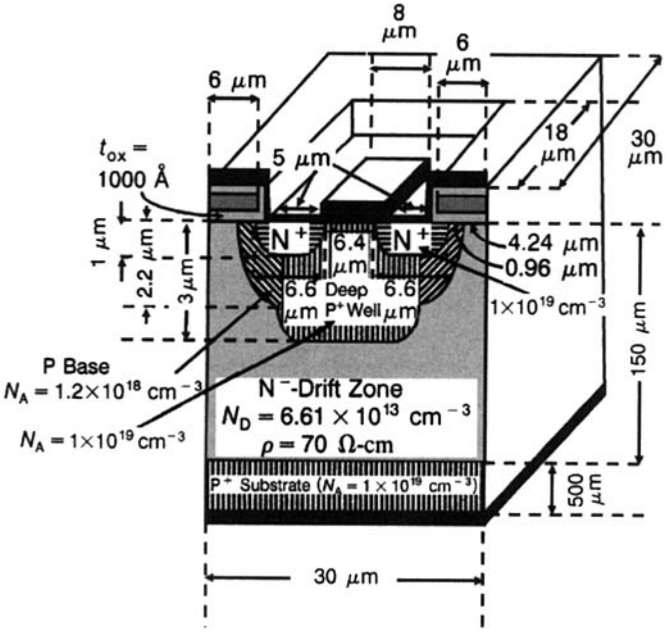

8 μm
6 μm
6 μm
5 μm
18 μm
30 μm
t_{ox} = 1000 Å
1 μm
2.2 μm
3 μm
N+
6.4 μm
N+
6.6 μm
Deep P+ Well
6.6 μm
4.24 μm
0.96 μm
1×10¹⁹ cm⁻³
150 μm
P Base
N_A = 1.2×10¹⁸ cm⁻³
N_A = 1×10¹⁹ cm⁻³
N⁻-Drift Zone
N_D = 6.61 × 10¹³ cm⁻³
ρ = 70 Ω-cm
P+ Substrate (N_A = 1 × 10¹⁹ cm⁻³)
500 μm
30 μm

Figure 7.9 Three-dimensional drawing of 0.96-μm channel-length high-voltage IGBT cell.

1 μm and that of the P-base region is 2.2 μm, leading to a channel length of 0.96 μm. The peak P-base doping is 1.2×10^{18} cm^{-3}. The IGBT half-cell had a deep P$^+$ region in the center of the P base, with a peak concentration of 1×10^{19} cm^{-3} to prevent latchup of the parasitic bipolar thyristor. The central P$^+$ region in the P base is 3 μm deep. Since the unit cell considered is a 30-μm $\times$ 30-μm square cell, the output current will be multiplied by 1.11×10^5 to give the current density (A/cm^2). This current density, when multiplied by the active area of IGBT chip, gives the total current capability of the particular device. The ambipolar lifetime in these simulations was 10 μsec. Three-dimensional pictures of the 2.4-μm high-voltage and medium-voltage IGBT cells are presented in Fig. 7.8 while a 3-D view of 0.96-μm high-voltage IGBT cell is shown in Fig. 7.9.

Figure 7.10 compares the forward characteristics of the 0.96-μm-channel-length high-voltage IGBT cell with the 2.4-μm-channel-length medium- and high-voltage cells at 300 K. It is readily seen that the forward voltage drop of a 0.96-μm cell is lowest (1.4 V) while the forward drop of a 2.4-μm medium-voltage cell (1.9 V) is smaller than that of its high-voltage counterpart (3.7 V). All these potential drops are measured at a collector–emitter current density of 100 A/cm^2, as marked in Fig. 7.10. Furthermore, as indicated in the diagram, saturated collector–emitter current density of a 0.96-μm-channel-length cell is 780 A/cm^2, that of a 2.4-μm-channel-length medium-voltage cell is 375 A/cm^2, and that of an equal-channel-length

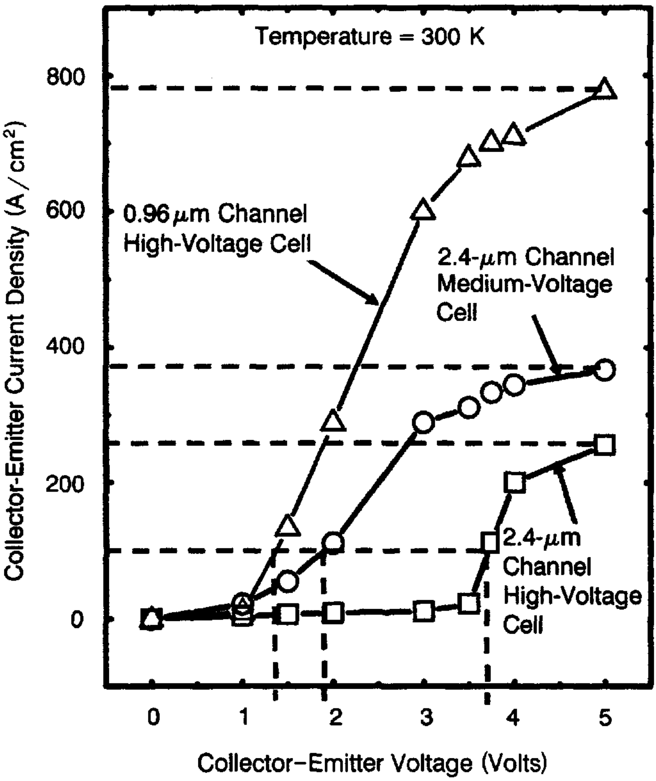


Figure 7.10 Comparison of steady-state forward conduction characteristics of a 2.4 μm-channel-length high-voltage IGBT cell, a 2.4-μm-channel-length medium-voltage IGBT cell, and a 0.96-μm-channel-length high-voltage IGBT cell, all simulated at room temperature (300 K).

high-voltage cell is 260 A/cm^2, decreasing in the same order as the forward drop. Thus, to realize low forward drop and high saturation collector–emitter current density IGBTs, ideally speaking, IGBT design calls for using the shortest possible channel length to achieve the lowest channel resistance. But the channel should be sufficiently long consistent with the depth of the P base dictated by punchthrough voltage requirement.

Effect of Channel Length on Latching Current Density. Unfortunately, for the required shorter channel length, the junction depth of the P base must be smaller. It has the adverse effect of increasing the lateral resistance of P base for the emitter portion not overlying the central deep P$^+$ region. Concurrently, the shallow P base leads to an increase in the current gain of the upper N-P-N transistor. Higher injection efficiency due to a smaller base

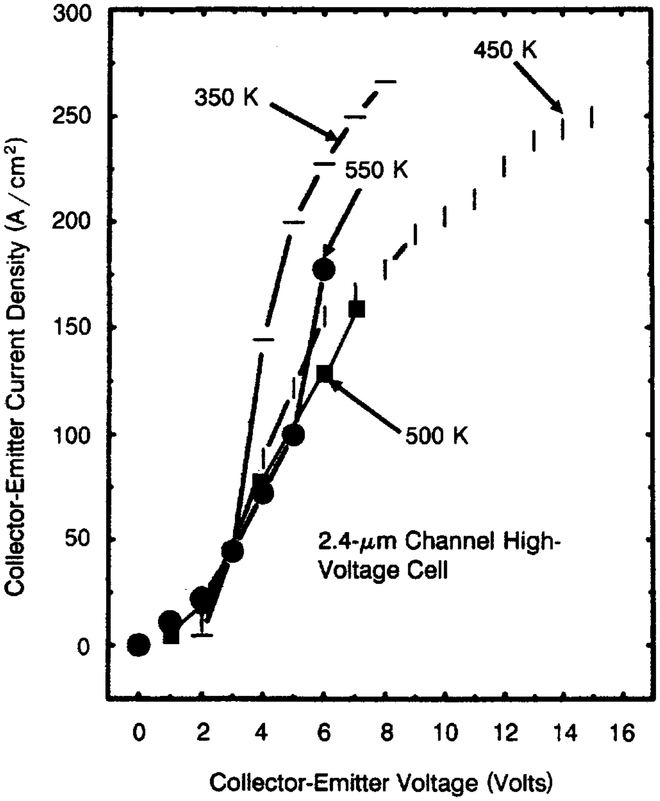


Figure 7.11 Simulated curves showing the dependence of forward characteristics of a 2.4-μm-channel-length high-voltage IGBT cell on temperature from 350–550 K.

charge, together with an increased base transport factor resulting from a narrow base width, is responsible for this current gain enhancement. Consequently, the IGBT latches at a lower current density, becoming thermally unstable. The device structure and profile must therefore be modified in such a way that latching occurs at elevated temperature. Performing a comprehensive thermal analysis therefore assesses device performance. Thermal stabilization can be achieved by various methods. Current–voltage characteristics of the resultant IGBTs have been generated and a comparison has been made in terms of their impact on device parameters and likely technological constraints during fabrication. It has been shown that safe IGBT design must take into account the thermal stability factor to achieve desired specifications.

Although the 0.96-μm-channel-length cell was superior to the 2.4-μm-channel-length IGBT cell due to its smaller forward drop, the former cell was found to be more prone to latching (Figs. 7.11–7.13). This was apparent from

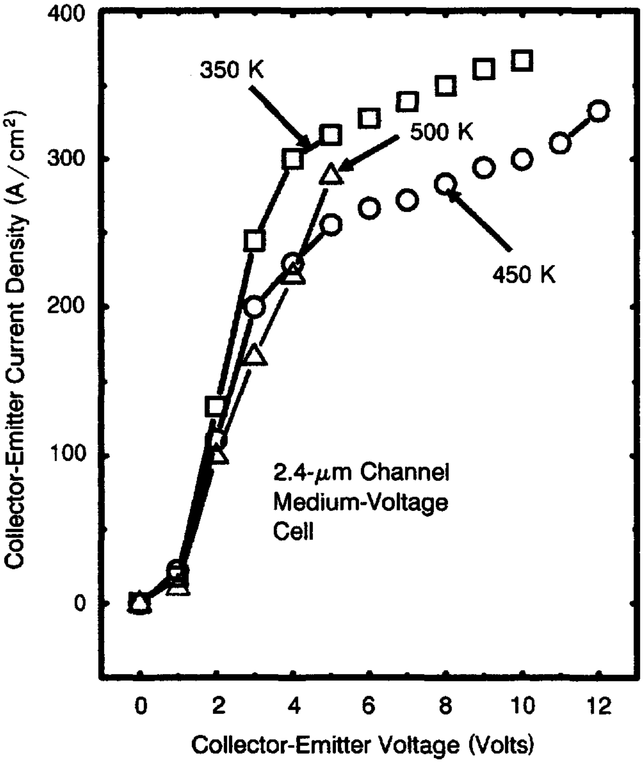


Figure 7.12 Simulated plots of forward characteristics of a 2.4-μm-channel-length medium-voltage IGBT cell as a function of temperature from 350–500 K.

the latching taking place at much lower current densities and also at lower temperatures for the 0.96-μm-channel-length cell as compared to 2.4-μm cell. Hence, a reduction of channel length to minimize forward voltage drop makes the IGBT susceptible to latchup. In other words, by decreasing the channel length, we have reduced the ON-state losses but the latchup problem has been aggravated. This vulnerability to latching will make the device susceptible to thermal runaway; thus during operation, one must keep away from the minimum latching temperature. Depending on the intended application, the designer must define adequate safety margins for junction temperature and current density to keep away from the minimum latching conditions.

To improve thermal ruggedness of the 0.96-μm-channel-length high-voltage IGBT cell, two approaches have been used. First, P-base concentration has been increased (Fig.7.14). This reduces the lateral resistance of the P base. The P-base concentration has been increased to 2×10^{18} cm^{-3}. The

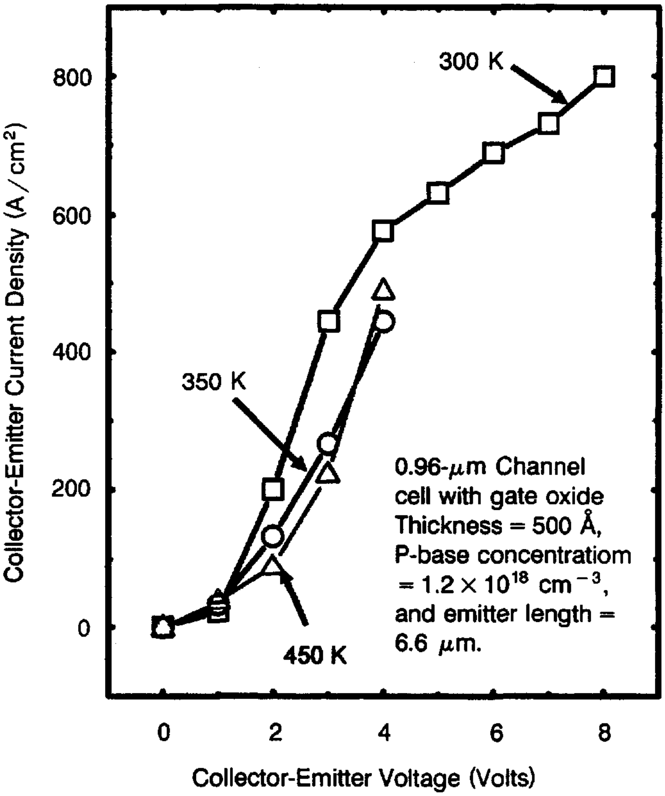


Figure 7.13 Simulated forward characteristics of a 0.96-μm-channel-length high-voltage IGBT cell at different temperatures in the range 350–500 K.

forward characteristics of this cell are more thermally stable. So, in IGBT design, one must seek a compromise between channel length and thermal stability for the required ON-state voltage. The threshold voltage and hence transconductance will not be affected if we could simultaneously decrease the gate oxide thickness in the correct ratio. So, an increase in P-base resistance with proportionate diminution in the gate oxide thickness to 500 Å provided an acceptable solution to the extent of ensuring safety of gate oxide from breakdown caused by spikes in the gate drive voltage.

Secondly, the emitter length has been decreased to 3.6 μm. Decrease in emitter length has been found to avert latchup up to 550 K while maintaining transconductance. This is illustrated in Fig. 7.15 for a 0.96-μm high-voltage cell. However, it requires finer geometries (2 μm) and thus imposes lithographic constraints.

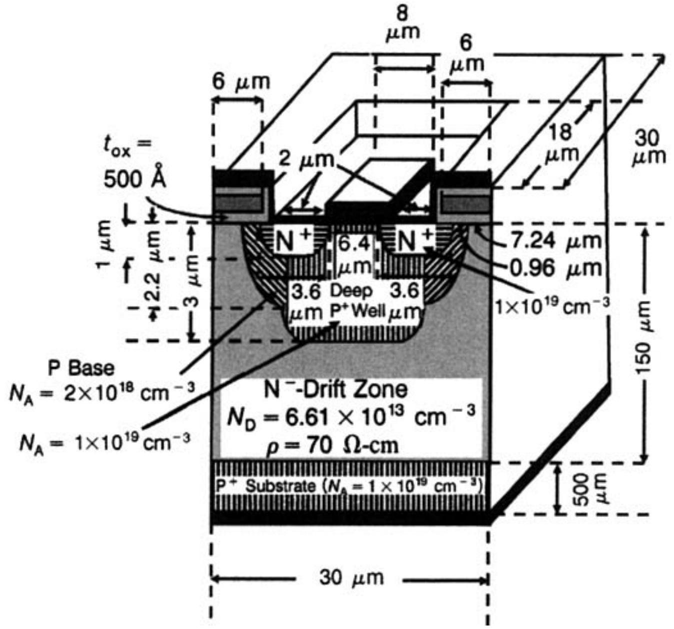

8 μm
6 μm
6 μm
t_{ox} = 500 Å
2 μm
18 μm
30 μm
1 μm
2.2 μm
3 μm
N+
6.4 μm
N+
3.6 μm
Deep P+ Well
3.6 μm
7.24 μm
0.96 μm
1×10¹⁹ cm⁻³
150 μm
P Base
N_A = 2×10¹⁸ cm⁻³
N_A = 1×10¹⁹ cm⁻³
N⁻-Drift Zone
N_D = 6.61 × 10¹³ cm⁻³
ρ = 70 Ω-cm
P⁺ Substrate (N_A = 1 × 10¹⁹ cm⁻³)
500 μm
30 μm

Figure 7.14 Three-dimensional drawing of a modified 0.96-μm-channel-length high-voltage IGBT cell.

Transfer, Blocking, and Transient Characteristics. The IGBT cell transfer characteristic (Fig. 7.16) does not saturate as does the DMOS cell, because the series resistance of the N epilayer in the IGBT is very low. This is due to conductivity modulation of the drift region, caused by hole injection from P^+ substrate. Figure 7.17 shows the blocking characteristics of the high-voltage IGBT cell at 300 K.

The P-I-N rectifier in the IGBT cell slows down its operation. The switching behavior of this P-I-N rectifier has been studied by applying a collector voltage pulse of magnitude 0.8 V and duration 1×10^{-2} μsec and observing the turn-off time (Fig. 7.18). The turn-off time (20 μsec) increases with temperature because, with rise of temperature, the carrier lifetime in the drift epilayer increases, decelerating the decay of carriers by recombination. So the injected carriers take a longer time to decay to their equilibrium values. Another reason is the consequent increase in the current gain, α_{PNP}, of the P-N-P transistor, due to temperature rise in the IGBT cell, resulting in the collector current in turn-off time being extended as the collector current has to decay from a higher initial value. This increase in the collector current turn-off time lengthens the turn-off process. Together these effects increase turn-off time of the IGBT cell with temperature.

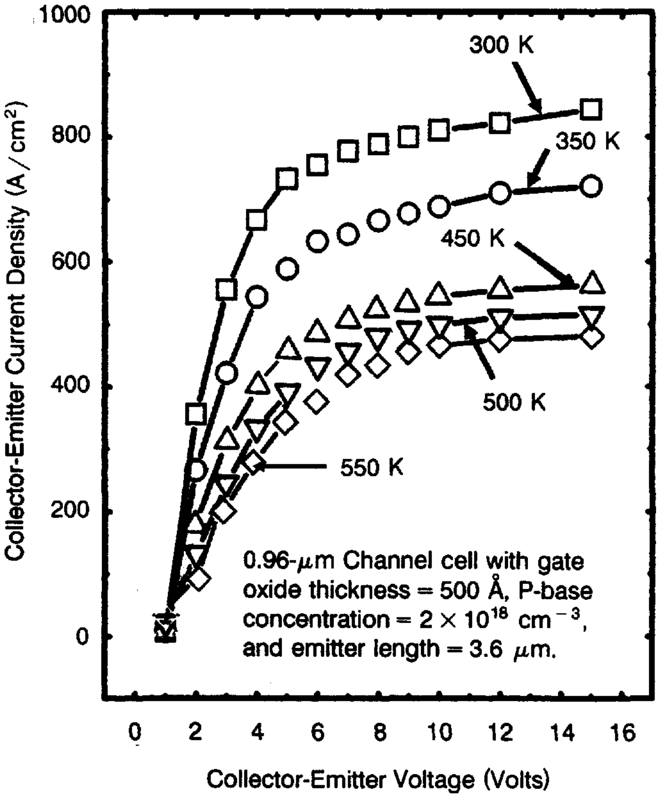


Figure 7.15 Temperature dependence of forward characteristics of a 0.96-μm-channel-length high-voltage IGBT cell with gate oxide thickness decreased to 500 Å, P-base concentration increased to 2×10^{18} cm^{-3}, and the emitter length reduced to 3.6 μm.

Reduction of MOS Capacitances in IGBT Design. During IGBT design and fabrication, special care is taken to restrict the MOS capacitances in the structure because they limit the high-frequency response. The main obstacle to high-frequency operation is the necessity to charge and discharge the input gate capacitance C_{iss} given by

$$C_{\mathrm{iss}}C_{\mathrm{GE}} + C_{\mathrm{M1}} \tag{7.21}$$

where C_{GE} is the gate–emitter capacitance and C_{M1} is the capacitance produced by the overlapping of the gate electrode over the N^--drift region. Capacitance C_{GE} consists of three components:

$$C_{\mathrm{GE}} = C_{\mathrm{N}^+} + C_{\mathrm{p}} + C_{\mathrm{M2}} \tag{7.22}$$

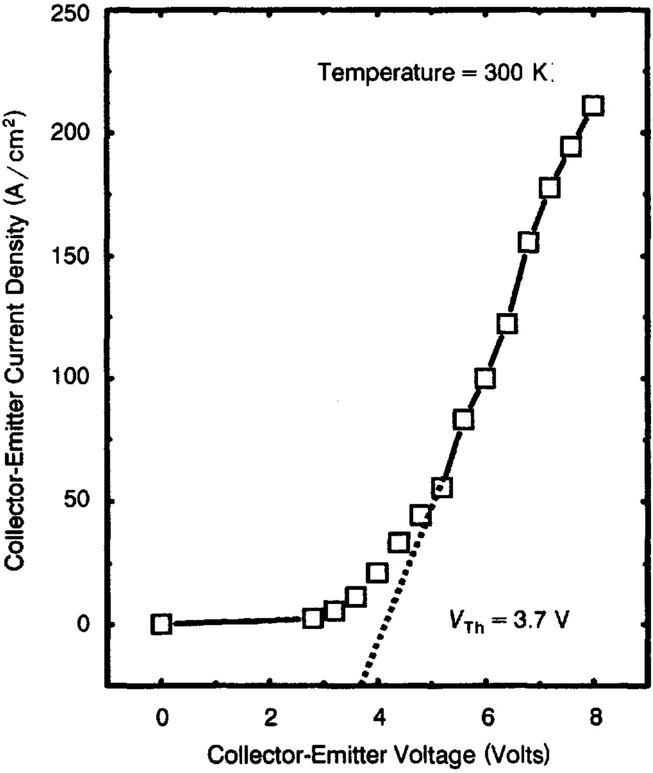


Figure 7.16 Transfer characteristic of a 2.4-μm-channel-length high-voltage IGBT cell.

In this equation, C_{N^+} is the capacitance produced by the overlapping of the gate electrode on the N^+ emitter. It is expressed as

$$C_{N^+} = \varepsilon_0 \varepsilon_{ox} A_{N+o} / t_{ox} \tag{7.23}$$

where ε_{ox} is the relative permittivity of SiO_2, t_{ox} is the gate oxide thickness, and A_{N+o} is the overlap area of the gate electrode over the N^+ emitter = Junction depth of N^+ emitter $(x_{N^+}) \times$ Channel width (Z). C_P arises from extension of the gate over the P base. It varies with gate bias and is decreased by using a small channel length. Capacitance C_{M2} originates from the extension of the emitter metal over the gate electrode. It is written as

$$C_{M2} = \varepsilon_0 \varepsilon_{ox} A'/t' \tag{7.24}$$

where A' is the overlapping area between emitter and gate electrodes and t'

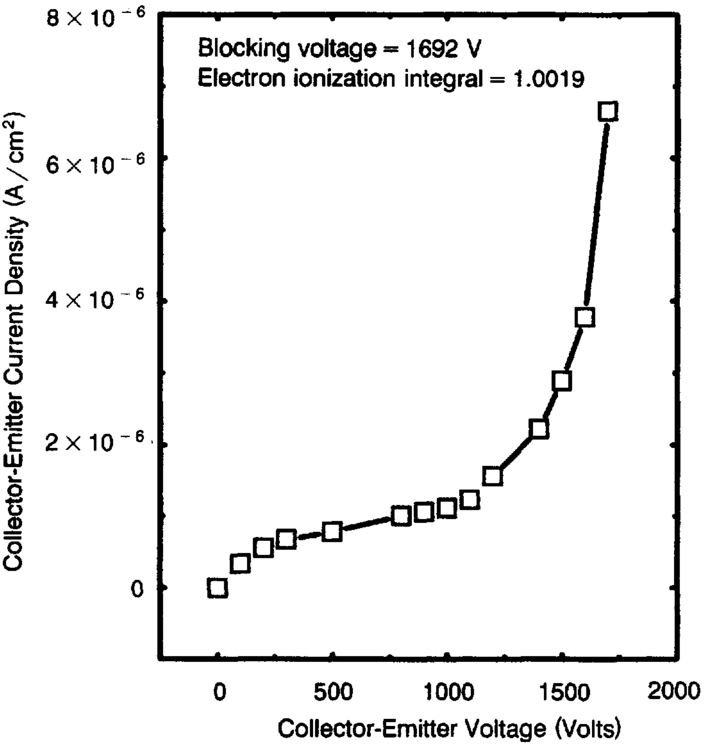


Figure 7.17 Blocking characteristic of a 2.4-μm-channel-length high-voltage IGBT cell at 300 K.

is the thickness of oxide layer covering the polySi gate. By increasing the thickness of this oxide layer, C_{M2} is minimized. C_{M2} is further decreased by confining the emitter metal to the diffusion window.

Looking back at Eq. (7.21), capacitance C_{M1} is given by

$$C_{M1} = (1 + g_m R_L) C_{GC} \tag{7.25}$$

representing the amplification of gate–collector capacitance C_{GC} by the Miller effect. g_m is the MOSFET transconductance and R_L is the load resistance connected between collector and emitter. By the Miller amplification effect, C_{GC} severely restrains the high-frequency response of IGBT. During ON-state of IGBT operation, the surface of the N^--drift region is in accumulation condition. Hence C_{GC} has a high value. During OFF state, with increasing collector bias, C_{GC} falls. C_{GC} is reduced by keeping the gate overlap over the N-drift region small.

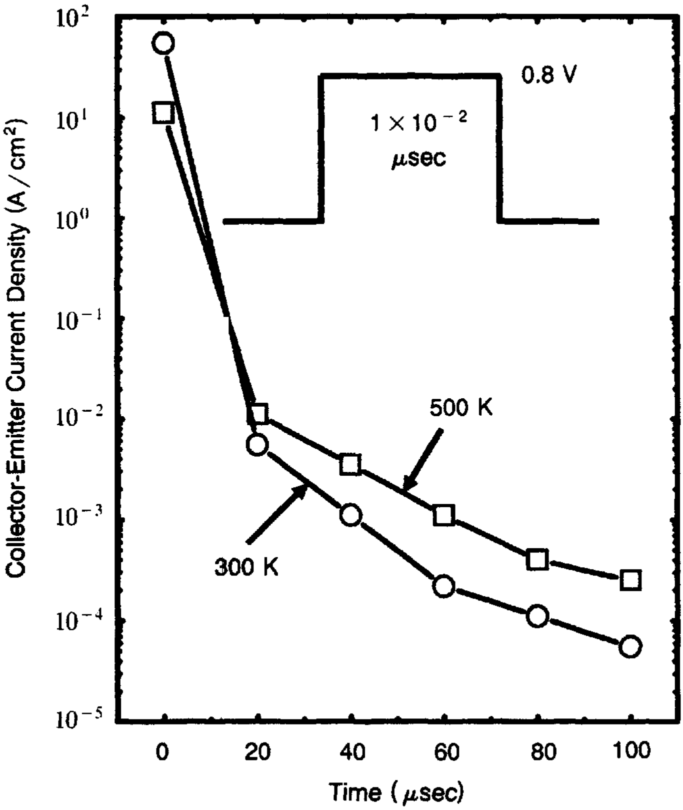


Figure 7.18 Transient characteristics of the PIN rectifier in a 2.4-μm-channel-length high-voltage IGBT cell at different temperatures. Applied collector voltage pulse is shown in the inset.

The storage component of turn-off time t_{off} of IGBT is a sensitive function of the feedback capacitance (C_{rss}) between the input and output terminals of IGBT as well as the distributed polysilicon resistance (R), which together determine the RC-time constant accounting for the delay. The delay time constant τ_d is derived by applying transmission line theory as [22] $\tau_d = 25\ L^2\ RC$, where L is the side of the square IGBT unit cell. The feedback capacitance is minimized by adopting a terraced gate structure in place of the conventional gate structure. This structure has been reported to decrease the feedback capacitance from 22 pF in the conventional gate to 2 pF in the terraced type at $V_{\text{CE}} = 2$ V. These two structures are compared in Fig. 7.19.

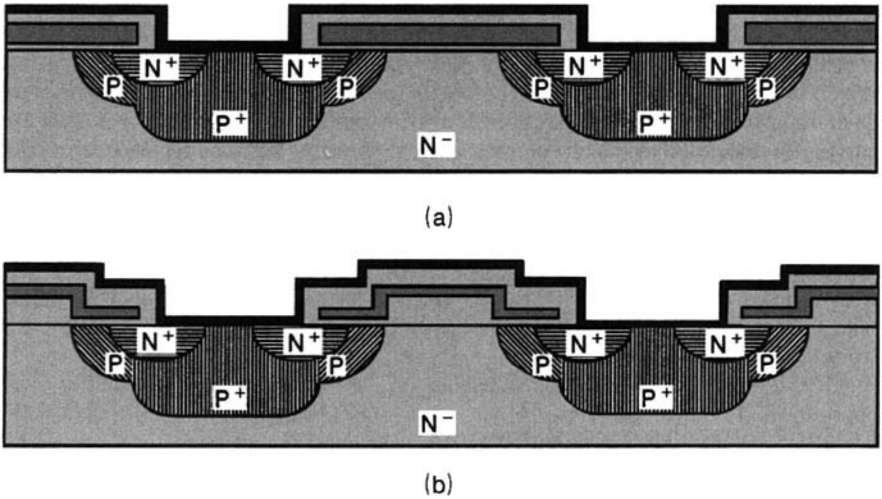


Figure 7.19 Comparison of the (a) conventional and (b) terraced gate structures.

7.3 OPTIMIZATION OF N-BUFFER LAYER STRUCTURE

The doping concentration and thickness of the N-buffer layer play a crucial role in determining the reverse breakdown voltage, forward voltage drop, saturated collector–emitter current, transconductance, latching current density, and turn-off time of the punchthrough IGBT [23]. It is natural to expect that the higher doping concentration of the N-buffer layer as compared to the N^--drift region lowers the reverse breakdown voltage by avalanche mechanism. The remaining parameters, as listed above, are degraded by the reduction in the current gain α_{PNP} of the PNP transistor associated with the fall in injection efficiency of the P^+ emitter. Based on the bipolar transistor–MOSFET model of the IGBT, the variation of these parameters with N-buffer layer concentration is investigated and exemplified in Example 7.3. The influence of doping concentration of N-buffer layer on current gain α_{PNP} is presented in Table E7.3.1 for the 0.96-μm IGBT cell, for three ambipolar lifetime values of 0.6, 1, and 10 μsec, taking the N-buffer layer thickness as 10 μm and the N^--drift region thickness as 90 μm. Because breakdown voltage and forward drop calculations have been extensively illustrated in the preceding, the effect of doping concentration of buffer layer on saturated collector–emitter current, transconductance, latching current density, and turn-off time of an IGBT cell are studied. The reader may perform the calculations for these parameters. Curves showing these variations are given in Figs. E7.3.1–E7.3.4.

Example 7.3 Consider the 0.96-μm-channel-length high-voltage IGBT cell with the modification that the N-drift region thickness is 100 μm, out of which 10 μm is consumed by the N-buffer layer. For three ambipolar lifetime values in the drift region (namely, 0.6, 1.0, and 10 μsec), calculate and tabulate the variation of current gain α_{PNP} of PNP transistor as a function of N-buffer layer concentration. Plot the curves for saturated collector–emitter current density, transconductance per cm^2, latching current density, and turn-off time with respect to the doping concentration of N-buffer layer in the range 1×10^{16} to 5×10^{18} cm^{-3}. Here a voltage of 2.5 V is dropped across the N-drift region/P-base junction during the operation of IGBT. Also, threshold voltage is 4 V and gate drive voltage is 10 V. You may take the inversion layer mobility to be 300 cm^2/V-sec. For latching current density calculation, take the length of the P base under the emitter as 2.6 μm and take that of the P$^+$ trough under the emitter as 4 μm.

From Eq. (4.62), the injection efficiency of a P$^+$ substrate for an N-buffer layer concentration of 1×10^{16} cm^{-3} is $\gamma = 1 - (1/12.4)(1 \times 10^{16}/1 \times 10^{19})(10/1) = 0.99919$. Similarly, the injection efficiency is calculated for different values of buffer layer concentration: 5×10^{16}, 1×10^{17}, 5×10^{17}, 1×10^{18}, 2×10^{18}, and 5×10^{18} cm^{-3}. Depletion layer thickness at 2.5 V is 7.06 μm so that the undepleted N-region width is $90 - 3 - 7.06 = 79.94$ μm. For ambipolar lifetime $\tau_a = 0.6$ μsec, ambipolar diffusion length $L_a = \sqrt{18.34 \times 0.6 \times 10^{-6}} = 33.17$ μm. Likewise, the diffusion lengths corresponding to $\tau_a = 1$ and 10 μsec are 42.83 and 135.43 μm, respectively. For $\tau_a = 0.6$ μsec, base transport factor $\alpha_T = 1/\{\cosh(79.94/33.17)\} = 0.1782$. In like manner, for $\tau_a = 1$ μsec, $\alpha_T = 0.30206$ and for $\tau_a = 10$ μsec, $\alpha_T = 0.8479$. Injection efficiencies γ and the current gains $\alpha = \gamma\alpha_T$ for the above buffer layer concentrations are calculated. The results are tabulated in Table E7.3.1.

Now $C_{ox} = 8.854 \times 10^{-14} \times 3.9/(500 \times 10^{-8}) = 6.906 \times 10^{-8}$ F/cm^2. Further, $\mu_{ns}C_{ox}Z/L_C = 300 \times 6.906 \times 10^{-8} \times 75.84 \times 10^{-4}/(0.96 \times 10^{-4}) = 1.6367 \times 10^{-3}$. And $I_{CE,sat} = \{\mu_{ns}C_{ox}Z/2L_C)\}(V_{GE} - V_{Th})^2\{1/(1 - \alpha_{PNP})$. Here $V_{GE} - V_{Th} = 10 - 4 = 6$ V. The saturation current values are calculated for buffer layer concentration by using the respective α_{PNP} value. These are converted into current densities by multiplying by 1.11×10^5, and the data are plotted in Fig. E7.3.1. In analogous fashion, transconductance g_{ms} is obtained from the relation $g_{ms} = (\mu_{ns}C_{ox}Z/L_C)(V_{GE} - V_{Th})\{1/(1 - \alpha_{PNP})\}$ for different buffer layer concentrations. The dependence of g_{ms} on buffer layer concentration is depicted in Fig. E7.3.2.

Table E7.3.1 Influence of Buffer Layer Concentration on Injection Efficiency and Current Gain

Serial No.	Buffer Layer Concentration (cm^{-3})	Injection Efficiency	Current Gain $\tau_p = 0.6$ μsec	Current Gain $\tau_p = 1.0$ μsec	Current Gain $\tau_p = 10$ μsec
1	1×10^{16}	0.9992	0.178	0.3018	0.847
2	5×10^{16}	0.996	0.1775	0.3008	0.8445
3	1×10^{17}	0.992	0.1768	0.2996	0.84115
4	5×10^{17}	0.96	0.1711	0.2899	0.814
5	1×10^{18}	0.92	0.164	0.2779	0.78
6	2×10^{18}	0.8387	0.1495	0.2533	0.711
7	5×10^{18}	0.597	0.1064	0.1803	0.5062

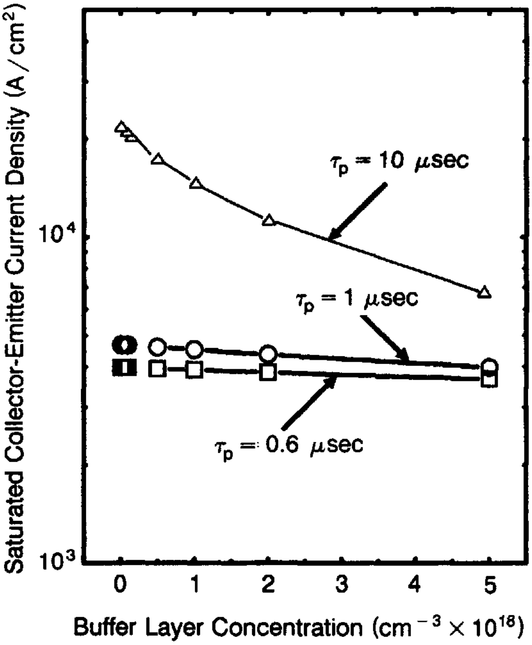


Figure E7.3.1 Influence of N-buffer layer concentration on saturated collector–emitter current density of IGBT for three carrier lifetimes (τ_p) of 0.6, 1, and 10 μsec.

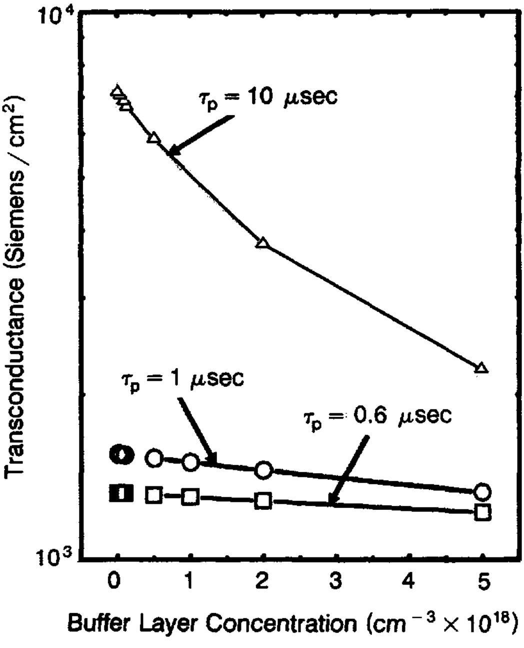


Figure E7.3.2 Variation of transconductance of IGBT with N-buffer layer concentration for three carrier lifetime values (τ_p) of 0.6, 1, and 10 μsec.

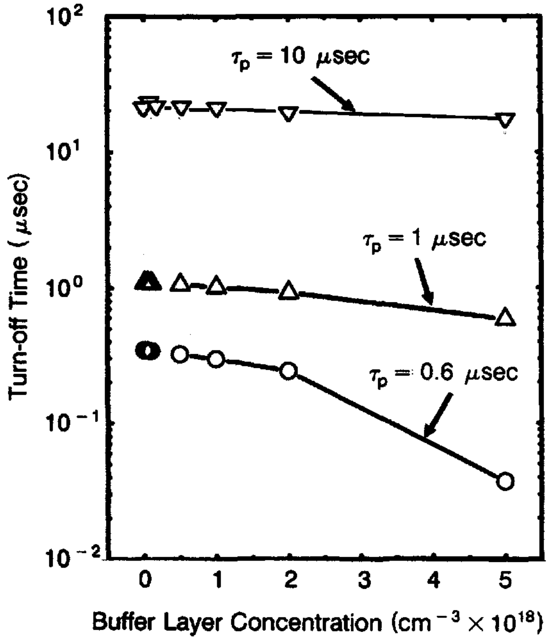


Figure E7.3.3 Dependence of turn-off time of IGBT on N-buffer layer concentration for three carrier lifetimes (τ_p) of 0.6, 1, and 10 μsec.

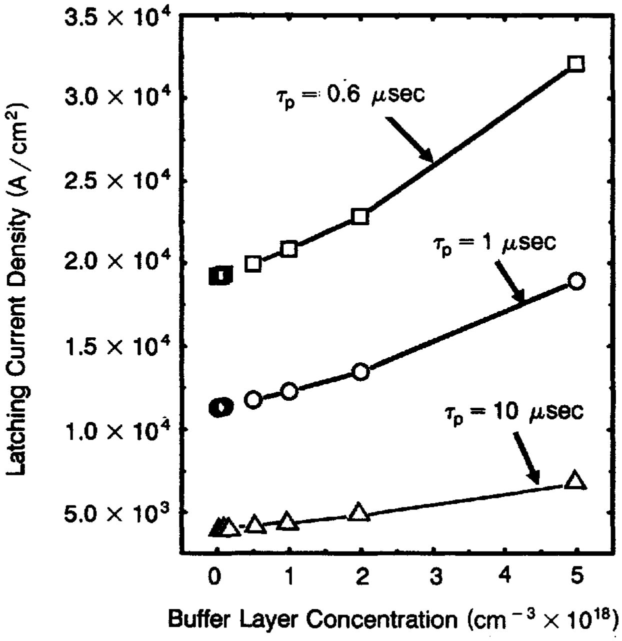


Figure E7.3.4 Impact of N-buffer layer concentration on latching current density of IGBT taking three minority-carrier lifetime values (τ_p) of 0.6, 1, and 10 μsec.

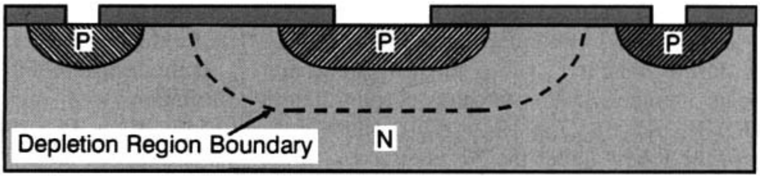
Figures E7.3.3 and E7.3.4 depict the variation of turn-off time and latching current density, respectively, with buffer layer doping density. The formula used for calculating the turn-off time is $t_{\text{off}} = \tau_{\text{HL}} \ln(10\ \alpha_{\text{PNP}})$, where τ_{HL} is the high-level lifetime = ambipolar lifetime τ_{a}. For latching current density calculations, the equation is $J_{\text{CL}} = V_{\text{bi}}/W_{\text{C}}\alpha_{\text{PNP}}(\rho_{\text{SB}}L_{\text{E1}} + \rho_{\text{SP+}}L_{\text{E2}})$, where half-cell width W_{C} is 15×10^{-4} cm, length of the P base under the N$^+$ emitter L_{E1} is 2.6 μm, and that of the P$^+$ region under the emitter L_{E2} is 4 μm. To calculate the sheet resistivity ρ_{SB} of the P$^-$ base under the emitter, the dopant concentration at a depth of 1 μm is determined for the Gaussian impurity diffusion profile by the formula $\exp(-x/x_j) \times$ Surface concentration $= \exp(-1/2.2) \times 1.2 \times 10^{18} = 7.617 \times 10^{17}\ \text{cm}^{-3}$. Similarly, for the sheet resistivity $\rho_{\text{SP+}}$ calculation at 1-μm depth we have dopant density $= \exp(-1/3) \times 1 \times 10^{19}\ \text{cm}^{-3} = 7.165 \times 10^{18}\ \text{cm}^{-3}$. Taking the hole mobilities in these regions as 100 cm^2/V-sec, we find that $\rho_{\text{SB}} = 683.792$ and $\rho_{\text{SP+}} = 43.613\ \Omega/\square$ respectively. Also V_{bi} = built-in potential between N$^+$ emitter ($1 \times 10^{19}\ \text{cm}^{-3}$) and P base ($1.2 \times 10^{18}\ \text{cm}^{-3}$) so that $V_{\text{bi}} = 0.999303$ V. Then, latching current density is obtained from the aforesaid equation for several current gain α_{PNP} values corresponding to the different buffer layer concentrations.

7.4 FIELD RING AND FIELD PLATE TERMINATION DESIGN

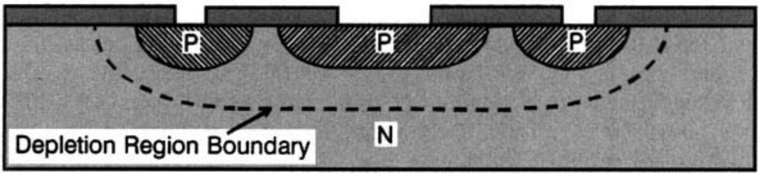
As we know, high-voltage planar PN junctions yield lower breakdown voltages under reverse bias than predicted by simple one-dimensional theory because of the three-dimensional effects of electric field crowding at the junction periphery. Consequently, specialized edge termination structures are indispensable in increasing the planar junction breakdown voltage for ideal bulk values. The floating field ring structure is attractive because it is formed simultaneously with the main junction, thus saving costly technological steps. It reduces the intensity of field crowding at the main junction by spreading the depletion layer past consequently lower potential floating junctions called *field rings*. Design optimization of the field ring structure calls for equal sharing of the electric field among the main junction and the field rings.

7.4.1 Critical Design Parameters

There are two critical design parameters of field rings, namely, field ring width and spacing. If the field ring is made too narrow, it will not affect the shape of the depletion region and its intended purpose will be lost. On the opposite extreme, too broad a field ring will unnecessarily waste silicon area. A simple design rule is to make the ring as wide as the depletion layer width. Similar remarks apply to the spacing between the main junction and the first field ring as well as ring-to-ring spacing. For example, if the separation between the main junction and the first ring is too large, the main junction will break down before its depletion region punches through to the first ring so that the ring will not be effective (Fig. 7.20a). On the contrasting side, if this spacing were too small for fear of breakdown, the full capability of the



(a)



(b)

Figure 7.20 Effect of spacing of the field ring on preventing voltage breakdown: (a) Large spacing and (b) small spacing.

ring for increasing the breakdown voltage will not be utilized (Fig.7.20b). More rings will be required to attain a given breakdown voltage rating.

First, the breakdown voltage of the main junction is calculated [24–28] from its junction depth, treating it as a cylindrical junction (Appendix 7.5):

$$\frac{(V_B)_{CYL}}{(V_B)_{PP}} = \frac{1}{2}\left\{\left(\frac{r_j}{W_c}\right)^2 + 2\left(\frac{r_j}{W_c}\right)^{6/7}\right\}\ln\left\{1 + 2\left(\frac{W_c}{r_j}\right)^{8/7}\right\} - \left(\frac{r_j}{W_c}\right)^{6/7} \tag{7.26}$$

where $(V_B)_{CYL}$ and $(V_B)_{PP}$ are the breakdown voltages of the cylindrical and the parallel-plane junctions respectively, r_j is the junction radius of curvature of the cylindrical junction, and W_c is the depletion layer width at breakdown. A useful empirical expression for the breakdown voltage of a cylindrical junction $(V_B)_{CYL}$ with respect to the parallel-plane junction $(V_B)_{PP}$ is

$$\frac{(V_B)_{CYL}}{(V_B)_{PP}} = \left\{0.871 + 0.125\ln\left(\frac{r_j}{W_{cPP}}\right)\right\}^2 \tag{7.27}$$

where W_c is the depletion layer width at breakdown for a parallel-plane junction given by

$$W_{cPP} = 2.67 \times 10^{10} N_D{}^{-7/8} \tag{7.28}$$

Referring to Fig. 7.21, taking the lateral diffusion to be equal to the vertical depth r_j, the mask length W_m (the distance of the floating field ring from the

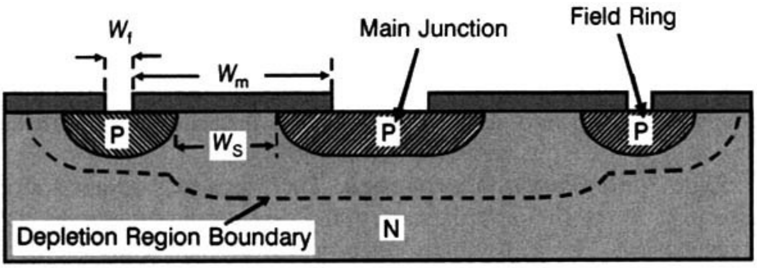

W_f
Main Junction
Field Ring
W_m
P
W_S
P
P
Depletion Region Boundary
N

Figure 7.21 Cross-sectional diagram of a diode with field ring showing the geometrical parameters and dimensions used in the analysis.

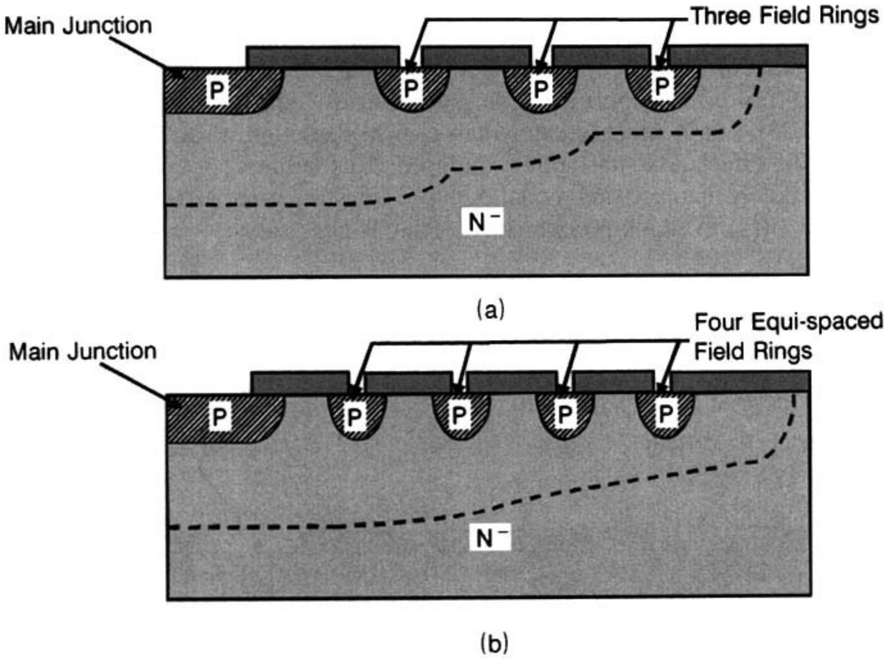

Main Junction
Three Field Rings
P
P
P
P
N⁻
(a)
Main Junction
Four Equi-spaced Field Rings
P
P
P
P
P
N⁻
(b)

Figure 7.22 Two approaches of field ring design: (a) Decreasing spacing and (b) uniform spacing.

main junction) is calculated [26] from the punchthrough voltage $(V_B)_{PT}$ by the equation

$$W_m = 2x_j + W_s = 2x_j + \sqrt{\frac{2\varepsilon_0 \varepsilon_s V_{PT}}{qN_D}} \tag{7.29}$$

where W_s is the spacing between the main junction and the first ring below the Si surface, taking into account the sideways diffusion from both sides.

7.4.2 ***Field Ring Design Approaches.*** There are two design approaches for field rings. In the first approach, both the ring width and spacing decrease as we recede away from the main junction (Fig. 7.22a). Thus the depletion layer is gradually extended as we move away from the main junction. Toward the device periphery, narrower field rings placed at shorter distance apart help in reducing the chip area. In the second approach, a larger number of narrow and evenly spaced rings are used (Fig. 7.22b). Less wide field rings placed close together allow the usage of more rings. A finer gradation of depletion layer is thus produced. Irrespective of the approach followed, generally three field rings suffice in controlling breakdown, but up to 10 rings have been used for achieving breakdown voltage approximating the parallel-plane junction. It may be mentioned here that for the field rings formed by deep diffusion of impurities, a large cylindrical junction radius (r_j) results (Fig. 7.23). For r_j much greater than critical depletion width W_{cPP}, the field crowding effects are substantially reduced. Furthermore, optimum ring spacing markedly increases and equally spaced rings are more effective [28]. If the ratio r_j/W_{cPP} is small, precise ring spacing is necessary.

Example 7.4 In an IGBT, the punchthrough voltage of the main junction to ring 1 is 225 V and that of ring 1 to ring 2 is 175 V. The N-base resistivity is 30 Ω-cm and the P-base junction depth is 3 μm.

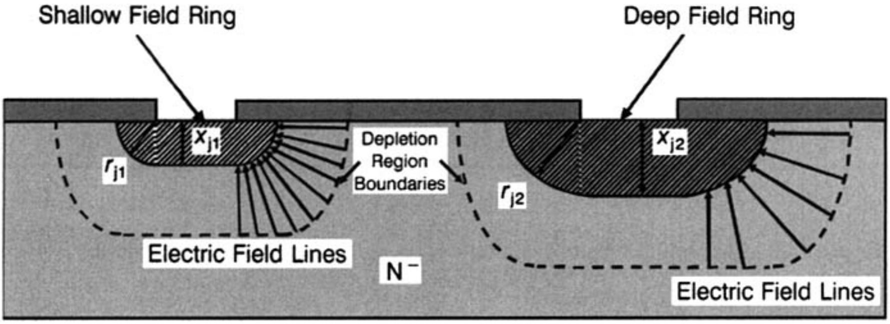


Figure 7.23 Effect of junction depth x_j on radius of curvature r_j and electric field crowding, $x_{j2} > x_{j1}$, $r_{j2} > r_{j1}$.

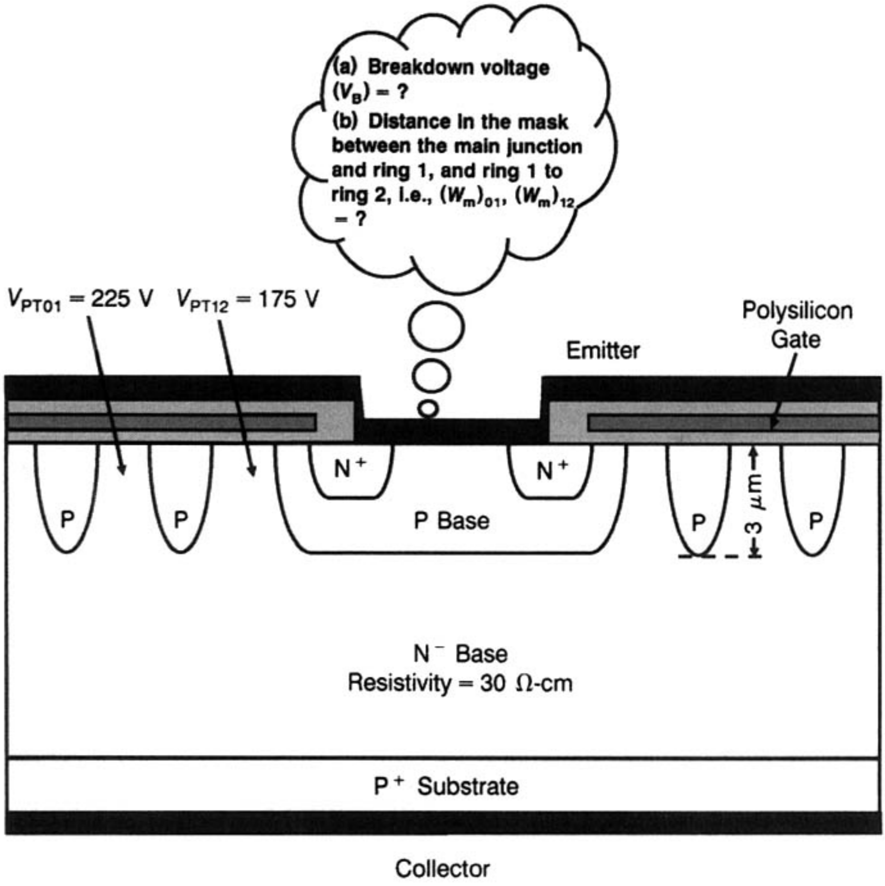

(a) Breakdown voltage (V_B) = ?
(b) Distance in the mask between the main junction and ring 1, and ring 1 to ring 2, i.e., $(W_m)_{01}$, $(W_m)_{12}$ = ?
V_{PT01} = 225 V
V_{PT12} = 175 V
Emitter
Polysilicon Gate
N+
N+
P Base
P
P
P
P
3 μm
N⁻ Base
Resistivity = 30 Ω-cm
P+ Substrate
Collector

Figure E7.4.1 IGBT structure for Example 7.4.

(a) What is the breakdown voltage of this two field ring structure?

(b) Calculate the distance in the mask between the main junction and ring 1, as well as between ring 1 and ring 2.

(a) The breakdown voltage of a structure with two floating field rings (Fig. E7.4.1) can be written as

$$V_B = V_{PT01} + V_{PT12} + V_{CL2} \tag{E7.4.1}$$

where V_{PT01} is the punchthrough voltage of the main junction to the first ring, V_{PT12} is that of the first ring to second ring, and V_{CL2} is the breakdown voltage of the second ring due to field crowding caused by junction curvature in a cylindrical junction. To apply Eq. (7.28), N_D is obtained from N$^-$-base resistivity (ρ) using the formula

$$N_D = 1/(\rho e \mu) \tag{E7.4.2}$$

where e denotes the electronic charge and μ denotes the electron mobility. This gives $N_D = 1/(30 \times 1.6 \times 10^{-19} \times 1350) = 1.54 \times 10^{14}$ cm^{-3}. Also, $r_j = 0.8 \times$ junction depth $= 0.8 \times 3 = 2.4$ μm $= 2.4 \times 10^{-4}$ cm. Now from Eq. (7.28), we have $W_{c,PP} = 2.67 \times 10^{10}(1.54 \times 10^{14})^{-0.875} = 0.01029$ cm. Therefore from Eq. (7.27), $(V_B)_{CYL2}/(V_B)_{PP} = \{0.871 + 0.125 \ln(r_j/W_{c,PP})\}^2 = \{0.871 + 0.125 \ln(2.4 \times 10^{-4}/0.01029)\}^2 = 0.16097$. But from Eq. (4.22), $(V_B)_{PP} = 60(E_g/1.1)^{3/2}(N_B/10^{16})^{-3/4} = 60(1.54 \times 10^{14}/10^{16})^{-0.75} 1372.49$ V. So, $(V_B)_{CYL2} = 0.7637 \times (V_B)_{PP} = 0.16097 \times 1372.49 = 220.933 \cong 221$V. Thus from Eq. (E7.3.1), the required breakdown voltage of the structure is $V_B = 225 + 175 + V_{CL2} = 400 + 221 = 621$ V.

(b) Mask spacing between the main junction and ring 1 is

$$\begin{aligned}(W_m)_{01} &= 2 \times 3 \times 10^{-4} \\ &\quad + \left\{2 \times 8.854 \times 10^{-14} \times 11.9 \times 225/(1.6 \times 10^{-19} \times 1.54 \times 10^{14})\right\}^{1/2} \\ &= 4.9866 \times 10^{-3} \text{ cm} \cong 50\ \mu\text{m}\end{aligned}$$

Similarly, the mask spacing from ring 1 to ring 2 is

$$\begin{aligned}(W_m)_{12} &= 2 \times 3 \times 10^{-4} \\ &\quad + \left\{2 \times 8.854 \times 10^{-14} \times 11.9 \times 175/(1.6 \times 10^{-19} \times 1.54 \times 10^{14})\right\}^{1/2} \\ &= 4.4686 \times 10^{-3} \text{ cm} \cong 45\ \mu\text{m}\end{aligned}$$

Example 7.5 In an IGBT, three field rings are placed at equal distances of 40 μm in the mask, the P-base junction depth is 4 μm, and N$^-$-base concentration is 2×10^{14} cm^{-3}. Calculate the breakdown voltage of the device.

Here $W_m = 25 \times 10^{-4}$ cm, and $x_j = 4 \times 10^{-4}$ cm. Equation (7.29) may be recast in the form

$$V_{PT} = \left(W_m - 2x_j\right)^2 qN_D/(2\varepsilon_0\varepsilon_s) \qquad \text{(E7.5.1)}$$

from which the punchthrough voltage of main junction to ring 1 is $V_{PT01} = (40 \times 10^{-4} - 2 \times 4 \times 10^{-4})^2 \times 1.6 \times 10^{-19} \times 2 \times 10^{14}/(2 \times 8.854 \times 10^{-14} \times 11.9) = 155.5$ V. Similarly, $V_{PT12} = V_{PT23} = 155.5$ V. Now, $W_{c,PP} = 2.67 \times 10^{10}(2 \times 10^{14})^{-0.875} = 8.1867 \times 10^{-3}$ cm, and the breakdown voltage of ring 3 due to cylindrical junction effect is $(V_B)_{CYL3}/(V_B)_{PP} = \{0.871 + 0.125 \ln(r_j/W_{c,PP})\}^2 = \{0.871 + 0.125 \ln(0.8 \times 4 \times 10^{-4}/8.1867 \times 10^{-3})\}^2 = 0.2169$. But from Eq. (4.22), $(V_B)_{PP} = 60(E_g/1.1)^{3/2}(N_B/10^{16})^{-3/4} = 60(2 \times 10^{14}/10^{16})^{-0.75} = 1128.181$ V. So, $(V_B)_{CYL3} = 0.2169 \times (V_B)_{PP} = 0.16097 \times 1128.181 = 181.603$ V. Assuming that the ring widths are sufficiently large to be effective in reducing the depletion layer curvature, breakdown voltage of the structure having three field rings is $V_B = V_{PT01} + V_{PT12} + V_{PT23} + (V_B)_{CYL3} = 155.5 \times 3 + 181.6 = 648.1$ V.

7.4.3 Numerical Simulation of Breakdown Voltage of PIN diode with Field Limiting Rings. To calculate the breakdown voltage of a diode with field rings, the linear Poisson model with depletion region logic is applied.

The boundary condition for the field rings is that net current flowing into each ring is zero. For satisfying the zero net current condition, the field ring junctions cannot be completely reverse- or forward-biased. The potential of each ring is equal to the lowest voltage V_{low} along its metallurgical ring–substrate junction minus the built-in potential (0.85 V). The algorithm is given as follows: (i) The lowest voltage $V_{\mathrm{low}(i)}$ along the metallurgical ring-substrate junction is calculated for each ring i. (ii) Looking through each node j, the electrostatic potential of this node $V_{\mathrm{new}(j)}$ is found by solving the Poisson's equation at that node. If node j is part of the ith ring and $V_{\mathrm{new}(j)}$ is $< V_{\mathrm{low}(i)} - 0.85$ V, then $V_{\mathrm{new}(j)}$ is clamped to the value $V_{\mathrm{low}(i)} - 0.85$ V. If node j is not part of a ring, then $V_{\mathrm{new}(j)}$ must be bounded above and below by anode and cathode potentials. (iii) When a sweep is over, the maximum potential change ΔV_{max} is compared with a predetermined change. If ΔV_{max} is sufficiently small, another sweep is not necessary.

After the electrostatic potential has been determined with the required accuracy, the hole ionization integrals are computed along the critical paths. Breakdown takes place when the hole ionization integral along one of the critical paths is unity.

7.4.4 Iterative Optimization of Ring Spacing

Suppose the spacing between the main junction and the first ring is W_{s1} and that between the first ring and the second ring is W_{s2}, and so on. In general, W_{si} is the separation between the $(i-1)$th and the ith rings, where the zeroth ring is the main junction. If the number of rings is fixed as n, a judicious choice of spacing between the rings maximizes the breakdown voltage. Any other spacing will decrease the breakdown voltage.

Consider a PIN diode with a single field ring. If the field ring is located far away from the main junction, the ionization integral at the main junction will be unity, and that at the field ring will be approximately zero at breakdown. On the contrary, if the field ring is too close to the main junction, the ionization integral will be zero at the main junction and that at the field ring will be unity at breakdown. For an optimally designed PIN diode with multiple field rings, the ionization integrals evaluated along each junction's critical path are simultaneously unity under breakdown condition. So, the differences between the ionization integrals of the adjacent rings guide us to determine the optimum spacing between the rings. The main junction is considered as the zeroth ring, and the field rings are numbered consecutively starting with the first ring. Therefore $II(0)$ is the ionization integral evaluated along the critical path near the main junction while $II(i)$ is the ionization integral evaluated along the critical path near the ith ring. Then $\Delta II(i) = II(i+1) - II(i) =$ difference of ionization integrals between the ith and $(i+1)$th rings. If $\Delta II(i)$ is positive, the rings are close together and the spacing should be increased. But if $\Delta II(i)$ is negative, the rings are far apart and should be brought in proximity.

The algorithm of the iteration method is: (i) An initial conjecture is made for the spacing between the rings. Then the PIN diode is simulated and the ionization integrals are computed for the main junction and field rings. (iii) Ring spacing is adjusted until the $\Delta II(i) = 0$ condition is achieved giving the optimum spacing.

7.4.5 Field Ring Design by Making Electric Field Profile Uniform in Quasi-Three-Dimensional Simulations

The optimization of field ring structure is extremely intricate because the results are a sensitive function of the solution method and grid conditioning [29]. For simulating the potential in the field ring, the electron quasi-Fermi potential is calculated for each ring involving a coupled solution of Poisson's equation and both the current continuity equations to achieve accuracy. Device gridding decides the stability of convergence. Suitable meshes are often obtained heuristically. A minimum mesh size of 0.1 μm near the floating junctions gives good convergence stability and accurate solutions. It must be further mentioned that 2-D simulations commonly employed for this work are found to grossly overestimate the breakdown voltage of planar junctions. For effective design and implementation of field ring termination, quasi-3-D simulations using MEDICI software package have been performed [2] in cylindrical symmetry to take 3-D electric field crowding effects into consideration.

In the approach described [2], first the main planar junction appended, with a single floating field ring, is analyzed. The breakdown voltage is computed for different spacings S_1 between the main junction and the first ring. On successively increasing the spacing from zero toward higher values, the breakdown voltage increases. Then it attains a maximum value, followed by a decline. Supposing that we begin from the maximum breakdown voltage obtained at the optimized spacing $S_1 = x$ and move toward either side. When the spacing is made more than x, the peak electric field is transferred to the edge of the main junction and the breakdown voltage is diminished. Similarly, when the spacing is decreased to a value smaller than x, the peak electric field is shifted to the ring edge but the shielding effect of the field ring on the main junction is reduced, resulting in a fall of breakdown voltage. Thus a uniform electric field profile must be maintained, both at the main junction and the field ring, in order that the full capability of the structure is utilized. After the main junction with the first ring has been optimized, another ring is added to the structure at a spacing S_2 from the first ring. It is understood that the electric field profile is completely transformed by the insertion of the second ring so that the previous field distribution no longer holds. This is because the spreading of the depletion layer now acquires a different shape. The peak electric field shifts toward the main junction and the first ring spacing must be decreased to make the field uniform. Proceeding in the same manner, more rings are added and the structure is optimized for each additional ring. Generally, the inner rings are brought closer for uniformity of electric field. In each case, the breakdown voltage obtained is calculated as a fraction of the ideal parallel-plane case. A final percentage of 85% completes the design. The optimized profile, with the different spacings of the field rings along with the ring width (say, 5 μm), is recorded.

7.4.6 Surface Charge Effects and Field Plate Attachment

Even the correct placement and width of field rings does not guarantee parallel-plane breakdown voltage because of the fixed positive charges in the oxide, a process-dependent parameter. Contamination of the Si surface by charged species during processing or introduction of charges by the operating stresses affects the subsurface field distribution and the nature of spreading of the depletion region, and it also alters the voltage at which punchthrough occurs from the main junction to the field ring or between any two consecutive rings. The nature and extent of the influence exerted by the surface charges on breakdown voltage is primarily determined by their magnitude and polarity. This positive charge attracts electrons toward the Si surface, tantamount to a local increase in doping density of the N^- base at the surface. As a result, the depletion layer is constricted near the surface (Fig. 7.24a). The depletion layer curvature produced in this manner, lowers the

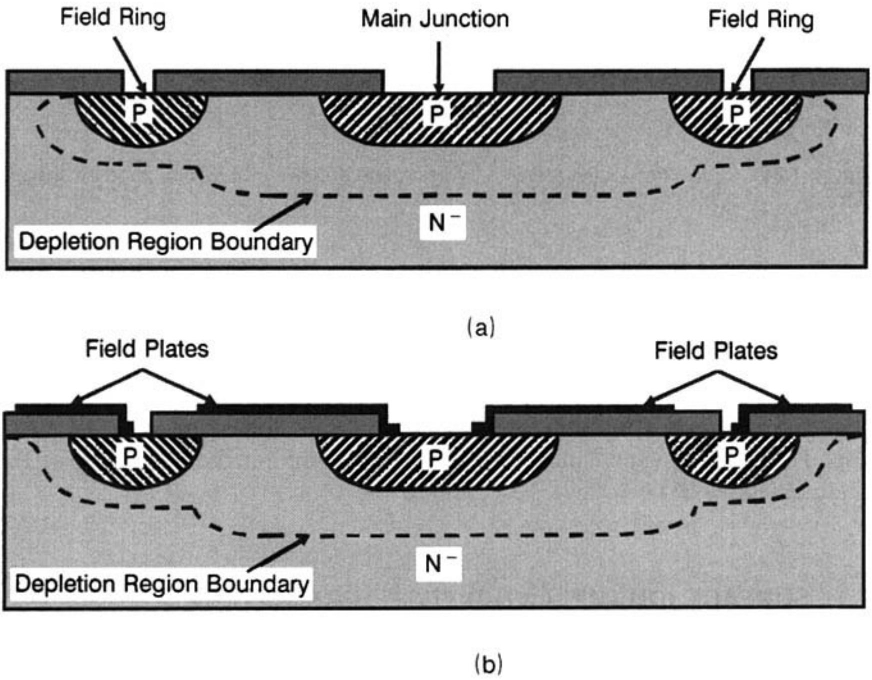
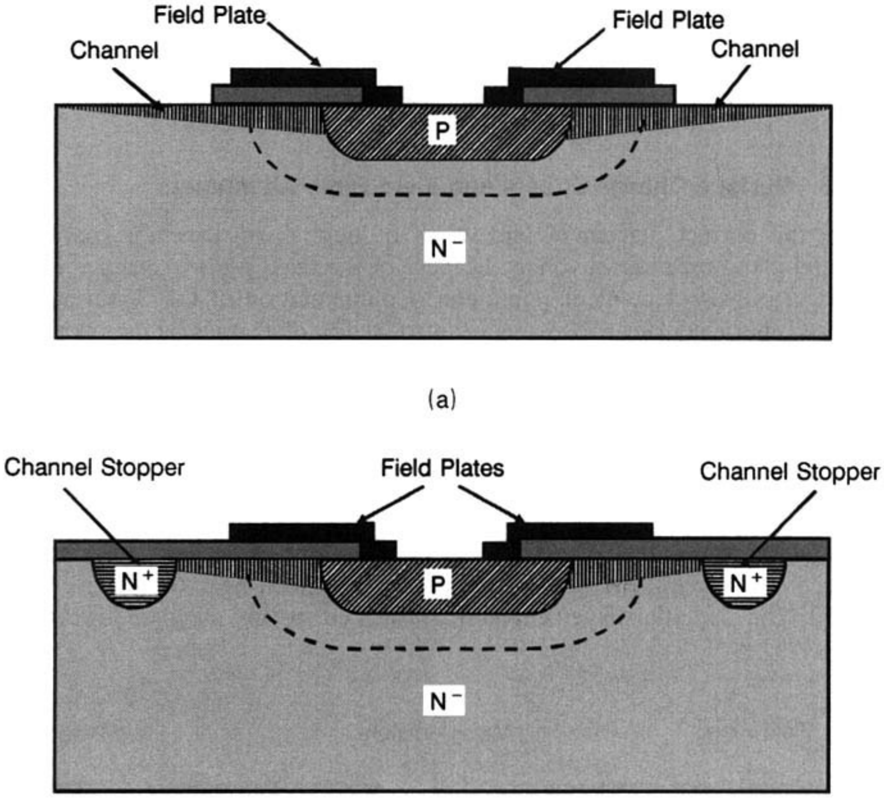


Figure 7.24 (a) Influence of surface-state charges on the depletion layer curvature. (b) Use of field plates to counter the effect of surface charges.



(a)

(b)

Figure 7.25 (a) Inversion layer creation by field plate potential. (b) Its stoppage by channel stopper.

breakdown voltage. The effect is counteracted with the help of a *field plate* [30], which is simply an extension of metallization over this higher electron concentration region (Fig. 7.24b). The negative potential applied to the field plate prevents such edge curvature of depletion region. But too high a negative bias may cause inversion, which is annulled by placing a heavily doped N^+ region called the *channel stopper* to inhibit the spreading of the channel (Fig. 7.25).

7.5 SURFACE ION-IMPLANTED EDGE TERMINATION

In this method, an accurately controlled quantity of charge is introduced on the surface near the junction at its edges by using ion implantation [31, 32], as shown in Fig. 7.26. The success of the method depends on the accuracy

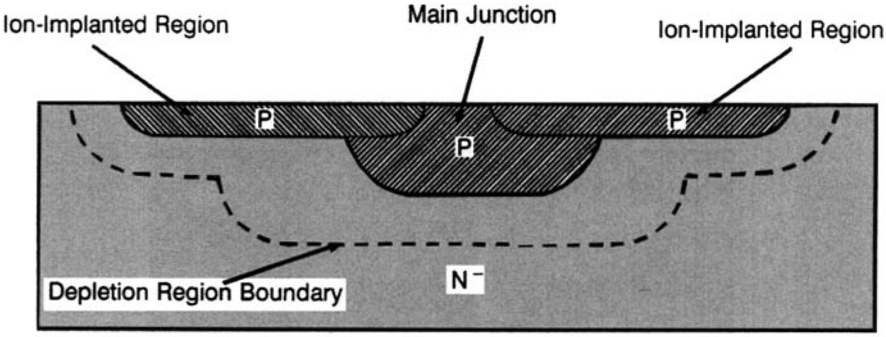


Figure 7.26 Schematic diagram of a planar diode with ion-implanted region as high-voltage termination.

achieved. Better than 1% accuracy is desirable. If the dose of the implanted charge is very low, the electric field distribution is imperceptibly influenced and the breakdown voltage remains unaffected. On the opposite side, too high a dose will extend the junction to the full length of the implanted region, if the latter is short in length. But the extended region has a very small radius of curvature and will break down by field crowding at the edge. In order to make it effective in enhancing breakdown voltage, it is necessary that:

(i) The implanted region becomes fully depleted on applying reverse bias. Then the applied voltage will be supported by the full length of the ion-implanted region. To calculate the charge Q in the depletion layer of width W, it may be noted that this charge Q is obtained from the critical electric field $E_c = 2 \times 10^5$ V/cm, by the equation $Q = qN_DW = \varepsilon_0 \varepsilon_s E_c = 1.3 \times 10^{12}$ per cm^{-3} where W is the width of the depletion layer.

(ii) Length of the ion-implanted region should be much greater than the depletion layer thickness in the bulk. Then only premature surface breakdown is averted.

Although ion implantation gives sufficient precision, interface state charges and segregation of the dopant into the oxide cause problems. These difficulties are removed by using multiple zones of ion-implanted regions.

7.6 REDUCED SURFACE ELECTRIC FIELD (RESURF) CONCEPT FOR BREAKDOWN VOLTAGE ENHANCEMENT IN LATERAL IGBT

The principle of the RESURF technique can be introduced with reference to the basic structure illustrated in Fig. 7.27 [33]. This basic structure consists of a high-resistivity P^--substrate with an epitaxial N^- layer grown upon it and laterally bounded by a P^+ diffused region. It thus comprises two diodes: a

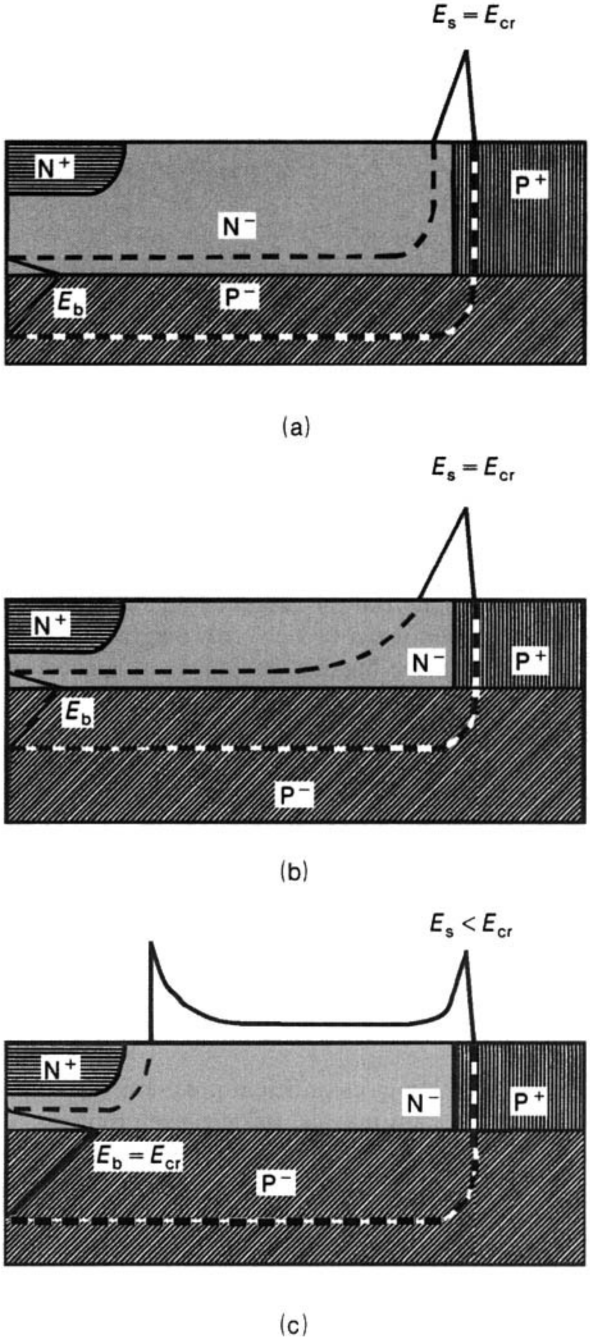

$E_s = E_{cr}$
N+
N−
P+
E_b
P−
(a)
$E_s = E_{cr}$
N+
N−
P+
E_b
P−
(b)
$E_s < E_{cr}$
N+
N−
P+
$E_b = E_{cr}$
P−
(c)

Figure 7.27 Electric field distribution in N$^-$ epitaxial layers of different thicknesses. (a) Thick N$^-$ expitaxial layer (Case I). (b) Thin N$^-$ expitaxial layer at low applied voltage (Case II). (c) Thin N$^-$ epitaxial layer at high applied voltage (Case III).

horizontal P^-N^- junction and a vertical P^+N^- junction. The breakdown voltage of the horizontal P^-N^- junction is high due to the high-resistivity P^- substrate used, while that of the vertical P^+N^- junction is low and is governed by the resistivity of the N^--layer. On applying a reverse bias to the N^+P^- junction, depletion regions formed by P^-N^- and P^+N^- junctions spread into the semiconductor. These depletion regions interact with each other, pushing the depletion region in the N^--drift region further and thereby reducing the electric field at the P^+N^- junction and the surface.

Three different cases are of interest:

Case I. For thick epitaxial layer (Fig. 7.27a), breakdown voltage is determined by the vertical P^+N^- junction because the depletion region at the surface is not affected by the horizontal P^-N^- junction.

Case II. For thin epitaxial layer (Fig. 7.27b), the depletion region of the vertical P^+N^- junction is reinforced by the horizontal P^-N^- junction. So, at a given applied reverse voltage, it extends over a long distance along the surface. Hence, surface electric field E_s is smaller than the critical electric field E_{cr}.

Case III. Below a certain epilayer thickness (Fig. 7.27c), the reduced surface electric field will remain smaller than E_{cr}, even at high voltages, eliminating the surface breakdown. Now horizontal junction P^-N^- decides the breakdown voltage, and ideal breakdown value is achieved. But due to the curvature of the N^+ contact, corner breakdown will occur at a voltage below the ideal voltage.

The above three cases demonstrate that to achieve ideal RESURF condition, with the widening of the depletion region, the electric field lines should terminate in the substrate and hence contribute only to the vertical component of the electric field. So, the breakdown will take place in the bulk instead of the surface. Proper RESURF calls for optimization of the concentration and thickness of the N^--drift region. With the help of two-dimensional numerical calculations, it has been shown that the electric field distribution at the surface will be symmetrical for optimum charge Q_D of the N^--drift region given by the product of its doping concentration N_D and thickness t_D:

$$Q_D = N_D t_D = 1 \times 10^{12}\ \text{cm}^{-2} \tag{7.30}$$

Thus by using high-ohmic substrates with thin epitaxial layers grown on them, and satisfying the criterion of Eq. (7.30), it is possible to make high-voltage devices in which the electric field distribution, and hence operation, differs from the conventional devices.

Figure 7.28 shows the cross section of a lateral IGBT based on the RESURF principle [34]. It consists of a horizontal P^+N^- junction along with

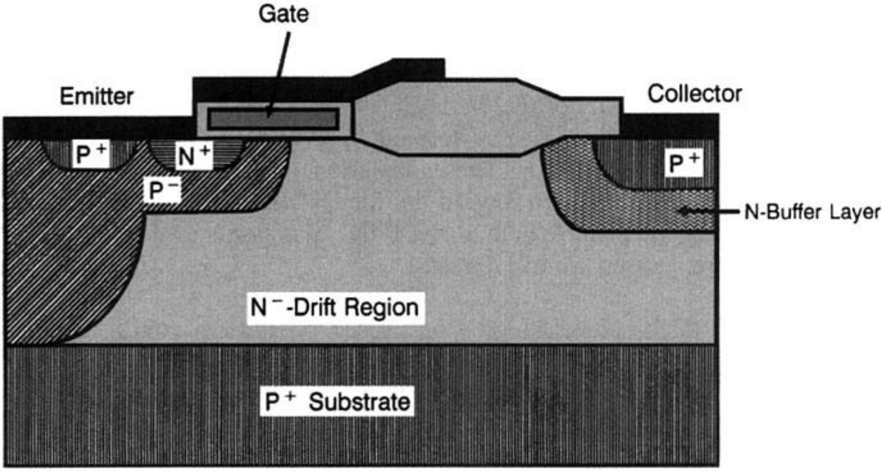


Figure 7.28 Lateral IGBT using the RESURF principle.

a vertical P^-N^- junction, with the concentration and thickness of the N^- layer in accordance with Eq. (7.30). So the electric field is distributed properly to provide the required blocking voltage.

7.7 CONCLUDING COMMENTS

Some general design guidelines can be recommended for achieving low forward voltage drop and high switching speed trade-off. *Reduction of forward drop of IGBT* requires that (i) the punchthrough structure should be used, (ii) minority-carrier lifetime should be high, (iii) the N^--epilayer should be thin, (iv) the channel resistance should be low necessitating thin gate oxides, and (v) the JFET resistance should be small through optimized cell spacing or trench–gate structures. For *switching speed enhancement* of IGBT, (i) the minority-carrier lifetime must be lowered (e.g., by electron irradiation) and (ii) the current gain of the PNP transistor must be small. The *safe operating area* (*SOA*) *capability* of an IGBT is restricted by the open-base SOA of its integral wide-base BJT. Additionally, multiple emitter pad layout decreases the current crowding in the vicinity of the emitter pad, thus preventing localized device failure due to local hot spots and electromigration effects.

IGBT design, like that of any power semiconductor device, seeks to achieve the best trade-off or compromise between the conflicting requirements of electrical and thermal specifications for the intended applications. It is impossible to design a single versatile device suitable for all applications. A person acquires the ability to design when he/she masters and practices the art of controlling a particular parameter and understands its impact on other parameters. A successful design engineer is one who is fully conversant with the ins and outs of this art and can easily maneuver the device properties.

REVIEW EXERCISES

7.1 Mention some advantages and shortcomings of using silicon as a fabrication material for IGBTs.

7.2 What special merits make silicon carbide a promising candidate material for high temperature power semiconductor devices?

7.3 What is meant by ionization coefficient of a carrier? How does the unity value of ionization coefficient signify the initiation of breakdown?

7.4 Comment on the statement "Design of IGBT seeks a compromise of forward conduction capability with each of the following parameters: blocking voltage, latching current density, and switching speed."

7.5 Two IGBTs A and B have carrier lifetimes of 5 μsec and 10 μsec in the N^--drift region. Qualitatively remark on their forward drops, latching currents, and turn-off times.

7.6 Why does the reduction of channel length make the IGBT more susceptible to latching? Explain.

7.7 Gate oxide thickness affects the threshold voltage, channel resistance, and transconductance of IGBT. Similarly, what parameters do the following influence: (i) P-base concentration and thickness, (ii) N^--drift region concentration and thickness, (iii) P^+ substrate concentration, (iv) Carrier lifetime in the N^--drift region, (iv) Buffer layer concentration, and (v) N^+ emitter concentration. Try to cite as many examples as you can, and thus relate the structural parameters of the device with its electrical parameters.

7.8 Explain, in the light of narrowing down emitter fingers to raise the latching current density of IGBT, that fabrication of thermally rugged IGBTs needs fine-resolution lithographic techniques.

7.9 Describe giving the relevant equations, the main physical models expressing the temperature dependence of physical properties of silicon such as intrinsic carrier concentration, mobility, carrier lifetime, and so on.

7.10 Explain in the context of latching immunization of power IGBT, how does thermal analysis aid in designing rugged IGBTs for high-temperature operation?

7.11 Briefly describe the special measures you will adopt for minimizing or eliminating the MOS capacitances that slow down IGBT operation.

7.12 Explain the criterion for choosing the width of the field ring? What will happen if the ring is excessively narrowed down? Point out the disadvantage of overly broadening the ring. What considerations enable you to decide the spacing between the main junction and the field ring or that between any two adjacent rings?

7.13 (a) How is the ring spacing optimized by numerical simulation approach? (b) What is the effect of oxide charge on the performance of the field ring termination?

7.14 How does the incorporation of field plate in the field ring structure help to raise the blocking voltage capability? Why is the need of a channel stopper felt?

7.15 Select the resistivity and thickness of the N^- base of an IGBT for a breakdown voltage of 500 V, given that the carrier lifetime in N^- base in the finished device is 1 μsec. Determine the punchthrough and avalanche breakdown voltages that will be obtained if the final carrier lifetime in the N^- base is 5 μsec? Calculate the avalanche breakdown voltage for the current gain at the specified 500-V value.

7.16 Recall the 0.96-μm-channel-length high-voltage IGBT cell, defined in Section 7.2.4 and illustrated in Fig. 7.7. Consider that the total N^--drift region thickness is 80 μm, out of which 6 μm is occupied by the N-buffer layer. Calculate and show in a table the change of current gain α_{PNP} of PNP transistor with the concentration of N-buffer layer.

7.17 An IGBT is designed with two field rings labeled as A and B. The punchthrough voltage of the main junction to ring A is 215 V, and that of ring A to ring B is 165 V. The resistivity of N^- base is 28 Ω-cm, and junction depth of the P base is 2.5 μm. Find the breakdown voltage of this field ring design. Calculate the distance on the mask between the main junction and ring A, and also that between ring A and ring B.

7.18 In an IGBT, four field rings are placed at equal distances of 30 μm in the mask. If the P-base junction depth is 3.7 μm and N^--base concentration is 3×10^{14} cm^{-3}, determine the breakdown voltage of the device.

REFERENCES

1. K. Chatty, S. Banerjee, T. P. Chow, and R. J. Gutmann, High-Voltage Lateral RESURF MOSFETs on 4H-SiC, *IEEE Electron Device Lett.*, Vol. 21, No. 7, July 2000, pp. 356–358.
2. D. C. Sheridan, G. Niu, J. N. Merrett, J. D. Cressler, C. Ellis, and C.-C. Tin, Design and Fabrication of Planar Guard Ring Termination for High-Voltage SiC Diodes, *Solid State Electron.*, Vol. 44, No. 8, August 2000, pp. 1367–1372.
3. A. Ohwada, H. Maeda, and T. Tanaka, Effect of the Crystal Orientation Upon Electron Mobility at the Si/SiO$_2$ Interface, *Jpn. J. Appl. Phys.*, Tokyo, Japan, Vol. 8, 1969, pp. 629–630.
4. O. Leistiko, A. S. Grove, and C. T .Sah, Electron and Hole Mobilities in Inversion Layers on Thermally Oxidized Silicon Surfaces, *IEEE Trans. Electron Devices*, Vol. ED-12, No. 5, May 1965, pp. 248–254.
5. J. Koga, S. Takagi, and A. Toriumi, A Comprehensive Study of MOSFET Electron Mobility in Both Weak and Strong Inversion Regimes, in *International Electron Devices Meeting, IEDM Technical Digest*, IEEE, New York, 1994, p. 475.

6. W. Fulop, Calculation of Avalanche Breakdown of Silicon P-N Junctions, *Solid State Electron.*, Vol. 10, No. 1, January 1967, pp. 39–43.
7. R. Van Overstraeten and H. DeMan, Measurements of the Ionisation Rates in Diffused Silicon P-N Junctions, *Solid State Electronics*, Vol. 13, No. 5, May 1970, pp. 583–608.
8. M. R. Pinto, C. S. Rafferty and R. W. Dutton, PISCES II: Poisson and Continuity Equation Solver, Technical Report, Stanford Electronics Laboratory, Department of Electrical Engineering, Stanford University, Palo Alto, 1984.
9. M. R. Pinto, C. S. Rafferty, H. R. Yeager, and R. W.Dutton, PISCES IIB: Supplementary Report, Technical Report, Stanford Electronics Laboratory, Department of Electrical Engineering, Stanford University, Palo Alto, 1985.
10. A. F. Franz, G. A. Franz, S. Selberherr, S. Ringhofer, and P. Markowich, Finite Boxes: A Generalization of the Finite Difference Method Suitable for Semiconductor Device Simulation, *IEEE Trans. Electron Devices*, Vol. ED-30, No.9, September 1983, pp. 1070–1082.
11. MEDICE, 2-D Semiconductor Device Simulator, Version 4.3, CA, Palo Alto, Avant Corp., 1999.
12. *DESSIS User's Guide*, ISE Integrated Systems Engineering AG, Zurich, Switzerland, 1994.
13. 1994–1998 Integrated Systems Engineering AG, Version 5.0, Zurich, Switzerland.
14. J. W. Slotboom and H. C. De Graff, Measurements of Bandgap Narrowing in Si Bipolar Transistor, *Solid State Electronics*, Vol. 19, No. 10, October 1976, pp. 857–862.
15. D. M. Caughey and R. F. Thomas, Carrier Mobilities in Silicon Empirically Related to Doping and Field, *Proc. IEEE*, Vol. 55, 1967, pp. 2192–2193.
16. G. Masetti, M. Severi, and S. Solmi, Modeling of Carrier Mobility Against Carrier Concentration in Arsenic-, Phosphorus-, and Boron-Doped Silicon, *IEEE Trans. Electron Devices*, Vol. ED-30, No.7, July 1983, pp. 764–769.
17. C. T. Sah, and H. C. Pao, The Effects of Fixed Bulk Charge on the Characteristics of Metal-Oxide-Semiconductor Transistors, *IEEE Trans. Electron Devices*, Vol. ED-13, No. 4, April 1966, pp. 393–409.
18. T. L. Chiu, and C. T. Sah, *Solid State Electronics*, Vol. 11, 1149, 1968.
19. K. M. Kramer, and W. N. G. Hitchon, *Semiconductor Devices: A Simulation Approach*, Prentice-Hall PTR, Englewood Cliffs, NJ, 1997.
20. V. K. Khanna, A. Kumar, B. Maj, A. Kostka, S. C. Sood, R. P. Gupta, and K. L. Jasuja, Physical Insight into Thermal Behaviour of Power DMOSFET and IGBT: A Two-Dimensional Computer Simulation Study, *Physica Status Solidi (a)*, Vol. 185, No. 2, June 2001, pp. 309–329.
21. V. K. Khanna, A. Kumar, B. Maj, A. Kostka, S. C. Sood, R. P. Gupta, and K. L. Jasuja, Power IGBT Design and Characterization by Two-Dimensional Thermal Simulation, *Proceedings of the XI-International Workshop on the Physics of Semiconductor Devices (IWPSD-2001)*, Dec. 11–15, 2001, SSPL, Delhi, Allied Publishers Ltd, New Delhi, pp. 436–439.
22. H. Yilmaz, K. O. Owyang, M. F. Chang, J. L. Benjamin, and W. R. V. Dell, Recent Advances in Insulated Gate Transistor Technology, *IEEE Trans. Ind. Appl.*, Vol. 26, No. 5, September/October 1990, pp. 831–834.
23. V. K. Khanna, A. Kumar, S. C. Sood, R. P. Gupta, K. L. Jasuja, B. Maj and A. Kostka, Investigation of Degeneracy of Current-Voltage Characteristics of Asymmetrical IGBT with N-Buffer Layer Concentration, *Solid State Electron.*, Vol. 45 , No. 10, October 2000, pp. 1859–1865.
24. Y. C. Kao and E. D. Wolley, High Voltage Planar P-N Junctions, *Proc. IEEE*, IEEE, Vol. 55, No. 8, August 1967, pp. 1409–1414.

25. V. A. K. Temple and M. S. Adler, Calculation of the Diffusion Curvature Related Avalanche Breakdown in High Voltage Planar P-N Junctions, *IEEE Trans. Electron Devices*, Vol. ED-22, No. 10, October 1975, pp. 910–916.
26. B. J. Baliga and S. K. Ghandhi, Analytical Solutions for the Breakdown Voltage of Abrupt Cylindrical and Spherical Junctions, *Solid State Electron.*, Vol. 19, No. 9, September 1976, pp. 739–744.
27. M. S. Adler, V. A. K. Temple, A. P. Ferro and R. C. Rustay, Theory and Breakdown Voltage for Planar Devices with a Single Field Limiting Ring, *IEEE Trans. Electron Devices*, Vol. ED-24, No. 2, February 1977, pp. 107–113.
28. D-. G. Bae and S-. K. Chung, An Analytical Model of Planar Junctions with Multiple Floating Field Limiting Rings, *Solid State Electron.*, Vol. 42, No. 3, March 1998, pp. 349–354.
29. V. Axelrad, Grid Quality and its Influence on Accuracy and Convergence in Device Simulation, *IEEE Trans. Comput.-Aided Des. Integr. Circuit Syst.*, Vol. 17, No. 2, February 1998, pp. 149–157.
30. F. Conti and M. Conti, Surface Breakdown in Silicon Planar Diodes Equipped with Field Plate, *Solid State Electron.*, Vol. 15, No. 1, January 1972, pp. 93–105.
31. A. S. Grove, O. Leistiko, and W. W. Hooper, Effect of Surface Fields on the Breakdown Voltage of Planar Silicon P-N Junctions, *IEEE Trans. Electron Devices*, Vol. ED-14, No. 3, March 1967, pp. 157–162.
32. V. A. K.Temple, Junction Termination Extension, A New Technique for Increasing Avalanche Breakdown Voltage and Controlling Surface Electric Fields in P-N Junctions, *International Electron Devices Meeting, IEDM Technical Digest*, IEEE, New York, Abstract 20.4, 1977, pp. 423–426.
33. J. A. Appels and H. M. J. Vaes, High Voltage Thin Layer Devices (RESURF Devices), International Electron Devices Meeting, Washington D.C., Dec 3–5, 1979, *Technical Digest*, IEEE, New York, pp. 238–241.
34. S. Hardikar, M. M. DeSouza and E. M. Sankara Narayanan, Advances in Devices and Technologies for Power and High Voltage Integrated Circuits, *Proceedings of the XI-International Workshop on the Physics of Semiconductor Devices (IWPSD-2001)*, Dec.11–15, 2001, SSPL, Delhi, Allied Publishers, New Delhi, pp. 422–431.

APPENDIX 7.1: MULTIPLICATION COEFFICIENT *M*

If the width of the depletion region is W, the multiplication coefficient or factor M for a carrier (electron or hole) is the ratio of the number of particles per unit volume leaving the space-charge region at $x = W$ to the number of particles per unit volume entering this region at $x = 0$. It is measured by the ratio of the current $I(W)$ at $x = W$ to the current $I(0)$ at $x = 0$. It is related to the ionization coefficient α (number of electron/hole pairs generated by a carrier while traversing unit distance through the depletion region in the direction of the field) as

$$M = \frac{1}{1 - \int_0^W \alpha \, dx} \tag{A7.1.1}$$

α varies with electric field and therefore with position in the space-charge region. As the width of the space-charge region changes with voltage, an approximate empirical relationship for M is useful:

$$M = \frac{1}{1 - \left(\frac{V_R}{BV}\right)^n}, \qquad 2 < n < 6 \tag{A7.1.2}$$

where $V_R < 0$ is the applied reverse voltage and BV is the breakdown voltage.

APPENDIX 7.2: V_{BR} EQUATION

V_{BR} is the collector–emitter breakdown voltage (BV_{CEO}) of a transistor with base open. On applying a negative voltage V_{CE} between the collector and emitter of a PNP transistor with base terminal floating, the base region acquires a potential intermediate between the emitter and collector potentials so that the emitter–base junction is slightly forward-biased. Then the collector current I_C = Reverse-biased generation current of the collector–base junction [which is the leakage current of the collector–base junction under open-emitter condition (I_{CBO})] + the current carried by injected carriers that reach the collector–base junction ($\gamma\alpha_T I_E$). Here, γ is the emitter injection efficiency, α_T is the base transport factor, and I_E is the emitter current. Since the current flowing through the emitter–base and collector–base junctions must be the same, we obtain

$$I_E = I_C = \gamma\alpha_T I_E + I_{CBO} \tag{A7.2.1}$$

or

$$I_{CEO} = \gamma\alpha_T I_{CEO} + I_{CBO} \tag{A7.2.2}$$

Near breakdown, the current flowing in a reverse-biased PN junction is amplified by the multiplication factor M due to avalanche so that Eq. (A7.2.2) is modified to

$$I_{CEO} = M(\gamma\alpha_T I_{CEO} + I_{CBO}) \tag{A7.2.3}$$

or

$$I_{CEO} = \frac{M I_{CBO}}{1 - \gamma\alpha_T M} \tag{A7.2.4}$$

Hence I_{CEO} will increase rapidly when $\gamma\alpha_T M = 1$ or, $M = 1/(\gamma\alpha_T)$. This is the breakdown condition.

Now the empirical formula for multiplication factor is

$$M = \frac{1}{1 - \left(\frac{BV_{CEO}}{BV_{CBO}}\right)^n} \tag{A7.2.5}$$

where BV_{CBO} is the breakdown voltage of the collector–base junction. Then,

$$M = \frac{1}{1 - \left(\dfrac{BV_{\mathrm{CEO}}}{BV_{\mathrm{CBO}}}\right)^{\mathrm{n}}} = \frac{1}{\gamma\alpha_{\mathrm{T}}} \quad \text{or} \quad 1 - \left(\frac{BV_{\mathrm{CEO}}}{BV_{\mathrm{CBO}}}\right)^{\mathrm{n}} = \gamma\alpha_{\mathrm{T}} \quad \text{or}$$

$$\left(\frac{BV_{\mathrm{CEO}}}{BV_{\mathrm{CBO}}}\right)^{\mathrm{n}} = 1 - \gamma\alpha_{\mathrm{T}} \tag{A7.2.6}$$

Therefore, BV_{CEO} (or V_{BR}) is

$$BV_{\mathrm{CEO}} = (1 - \gamma\alpha_{\mathrm{T}})^{1/\mathrm{n}} \cong (1 - \alpha_{\mathrm{T}})^{1/\mathrm{n}} \tag{A7.2.7}$$

if $\gamma = 1$.

APPENDIX 7.3: AVALANCHE BREAKDOWN VOLTAGE (V_{B})

Avalanche breakdown voltage of a PN junction is defined as the voltage at which the avalanche multiplication coefficient M becomes infinitely large. This condition occurs when the ionization integral (in approximate form)

$$\int_0^W \alpha\, dx = 1 \tag{A7.3.1}$$

where W is the width of the space-charge region and α is the ionization rate which varies with electric field as

$$\alpha = A \exp\left(-\frac{b}{E}\right)^{m} \tag{A7.3.2}$$

For electrons in Si, $A = 3.8 \times 10^6$ cm^{-1}, and $b = 1.75 \times 10^6$ V/cm; for holes in Si, $A = 2.25 \times 10^7$ cm^{-1}, and $b = 3.26 \times 10^6$ V/cm; and for both electrons and holes in Si, $m = 1$.

For an abrupt $\mathrm{P^+N}$ junction, Poisson's equation in the N region is

$$\frac{dE}{dx} = -\frac{qN_{\mathrm{B}}}{\varepsilon_0 \varepsilon_{\mathrm{s}}} \tag{A7.3.3}$$

where E is the electric field, q is the electronic charge, N_{B} is the background doping of the semiconductor ($\mathrm{N^-}$ region), ε_0 is the permittivity of free space, and ε_{s} is the relative permittivity of Si.

Integrating Eq. (A7.3.3) and applying the boundary condition that $E = 0$ at the edge of the depletion region (i.e., at $x = W$), we get

$$E(x) = \frac{qN_B}{\varepsilon_0 \varepsilon_s}(W - x) \tag{A7.3.4}$$

If V_a is the applied voltage, depletion region thickness is

$$W = \sqrt{\frac{2\varepsilon_0 \varepsilon_s V_a}{qN_B}} \tag{A7.3.5}$$

Now, maximum electric field E_m is obtained by putting $x = 0$ in Eq. (A7.3.4) because maximum field occurs at the junction. Applying this condition, we obtain with the help of Eq. (A7.3.5)

$$E_m = \sqrt{\frac{2qN_B V_a}{\varepsilon_0 \varepsilon_s}} \tag{A7.3.6}$$

Noting that V_a = avalanche breakdown voltage V_B, from Eq. (A7.3.6), we have

$$V_B = \frac{\varepsilon_0 \varepsilon_s {E_m}^2}{2qN_B} \tag{A7.3.7}$$

Putting $\varepsilon_0 = 8.854 \times 10^{-14}$ F/cm, $q = 1.6 \times 10^{-19}$ C; for Si, $\varepsilon_s = 11.9$, $E_m = 3 \times 10^5$ V/cm, Eq. (A7.3.7) becomes

$$V_B = 2.9633 \times 10^{17}\left(\frac{1}{N_B}\right) \tag{A7.3.8}$$

Due to the strong dependence of α on E, the maximum electric field varies gradually with background doping. As a first approximation, it is assumed that E_m has a fixed value for a particular semiconductor. Then V_B is approximately proportional to ${N_B}^{-1}$. Also, for a given background doping, the breakdown voltage increases with forbidden energy gap E_g of the semiconductor. Sze and Gibbons [1] formulated an approximate universal expression for abrupt junctions, valid for different semiconductors (Si, Ge, GaAs, and GaP) studied by them:

$$V_B = 60\left(\frac{E_g}{1.1}\right)^{3/2}\left(\frac{N_B}{10^{16}}\right)^{-3/4} \text{V} \tag{A7.3.9}$$

They computed the avalanche breakdown voltage numerically for PN junctions in the above semiconductors to obtain this general expression.

REFERENCES

1. S. M. Sze and G. Gibbons, Avalanche Breakdown Voltages of Abrupt and Linearly Graded P-N Junctions in Ge, Si, GaAs and GaP, *Appl. Phys. Lett.*, Vol. 8, March 1966, pp. 111–113.

APPENDIX 7.4: PUNCHTHROUGH VOLTAGE (V_{PT})

For an abrupt P^+N junction, the depletion layer widening is confined to N side of the junction only. The depletion layer width is given by

$$W = \left(\frac{2\varepsilon_0 \varepsilon_s V}{qN_D} \right)^{1/2} \tag{A7.4.1}$$

where V = applied reverse voltage + built-in potential (~ 0.7 V). Suppose the width of the N side is W_N, and the depletion layer spreads on this side such that it is one diffusion length short of the metallic contact, that is, the depletion layer penetrates up to a distance $(W_N - L_p)$ on the N side. Then the necessary voltage producing this spreading is the punchthrough voltage (V_{PT}). The diffusion length L_p is to be subtracted from the width of the N side because it is the distance that a carrier can travel without undergoing recombination. So, punchthrough will be deemed to be accomplished as soon as the depletion layer penetrates up to a distance $(W_N - L_p)$ on the N side. Therefore, replacing W by $(W_N - L_p)$ and V by V_{PT} in Eq. (A7.4.1), we get

$$W_N - L_p = \left(\frac{2\varepsilon_0 \varepsilon_s V_{PT}}{qN_D} \right)^{1/2} \tag{A7.4.2}$$

Squaring both sides, we have

$$(W_N - L_p)^2 = \frac{2\varepsilon_0 \varepsilon_s V_{PT}}{qN_D} \tag{A7.4.3}$$

or

$$V_{PT} = \frac{(W_N - L_p)^2 qN_D}{2\varepsilon_0 \varepsilon_s} \tag{A7.4.4}$$

APPENDIX 7.5: BV_{CYL}/BV_{PP} EQUATION

The potential distribution $V(r)$ in a cylindrical junction is expressed as [1, 2]

$$V(r) = \frac{qN_D}{2\varepsilon_0 \varepsilon_s} \left\{ \left(\frac{r_j^2 - r^2}{2} \right) + r_d^2 \ln\left(\frac{r}{r_j} \right) \right\} \tag{A7.5.1}$$

where N_{D} is the doping concentration of the N side, ε_0 is the prermittivity of free space, ε_0 is the relative permittivity of silicon, r_{j} is radius of curvature of the cylindrical junction, r_{d} is the radius of curvature of the depletion region, and r is radial distance. At breakdown $r =$ depletion region width at breakdown $= r_{\mathrm{d}}$; therefore, breakdown voltage of the cylindrical junction is

$$BV_{\mathrm{CYL}} = V(r_{\mathrm{d}}) = \frac{qN_{\mathrm{D}}}{2\varepsilon_0\varepsilon_{\mathrm{s}}}\left\{\left(\frac{r_{\mathrm{j}}^2 - r_{\mathrm{d}}^2}{2}\right) + r_{\mathrm{d}}^2 \ln\left(\frac{r_{\mathrm{d}}}{r_{\mathrm{j}}}\right)\right\} \tag{A7.5.2}$$

Also, at breakdown, electric field $E = E_{\mathrm{pC}} =$ peak electric field for cylindrical junction [2]:

$$E_{\mathrm{pCYL}} = \frac{qN_{\mathrm{D}}}{2\varepsilon_0\varepsilon_{\mathrm{s}}}\left(\frac{r_{\mathrm{d}}^2 - r_{\mathrm{j}}^2}{r_j}\right) \tag{A7.5.3}$$

or,

$$N_{\mathrm{D}} = \frac{2\varepsilon_0\varepsilon_{\mathrm{s}}E_{\mathrm{pCYL}}\,r_{\mathrm{j}}}{q\left(r_{\mathrm{d}}^2 - r_{\mathrm{j}}^2\right)} \tag{A7.5.4}$$

Substituting for N_{D} from Eq. (A7.5.4) into the first term of Eq. (A7.5.2) we get

$$\text{First term} = \frac{qN_{\mathrm{D}}}{2\varepsilon_0\varepsilon_{\mathrm{s}}}\left(\frac{r_{\mathrm{j}}^2 - r_{\mathrm{d}}^2}{2}\right) \times \frac{2\varepsilon_0\varepsilon_{\mathrm{s}}E_{\mathrm{pCYL}}\,r_{\mathrm{j}}}{q\left(r_{\mathrm{d}}^2 - r_{\mathrm{j}}^2\right)} = -\frac{E_{\mathrm{pCYL}}\,r_{\mathrm{j}}}{2} \tag{A7.5.5}$$

Since [2]

$$E_{\mathrm{pCYL}} = \left(\frac{3.25 \times 10^{35}}{r_{\mathrm{j}}}\right)^{1/7} \tag{A7.5.6}$$

Eq. (A7.5.5) becomes

$$\begin{aligned}\text{First term} &= -\left(\frac{3.25 \times 10^{35}}{r_{\mathrm{j}}}\right)^{1/7} \times \frac{r_{\mathrm{j}}}{2} = \frac{(3.25 \times 10^{35})^{1/7}}{2} \times r_{\mathrm{j}}^{1-1/7} \\ &= -\frac{(3.25 \times 10^{35})^{1/7}}{2} \times r_{\mathrm{j}}^{6/7}\end{aligned} \tag{7.5.7}$$

Avalanche breakdown voltage for parallel-plane junction is [2]

$$BV_{\mathrm{PP}} = 5.34 \times 10^{13} N_{\mathrm{D}}^{-3/4} \tag{A7.5.8}$$

Critical depletion region width at the point of breakdown is

$$W_c = 2.67 \times 10^{10} N_D^{-7/8} \tag{A7.5.9}$$

giving

$$N_D^{-7/8} = \frac{W_c}{2.67 \times 10^{10}} \quad \text{or} \quad N_D^{-1} = \left(\frac{W_c}{2.67 \times 10^{10}}\right)^{8/7}$$

Therefore,

$$N_D^{-3/4} = \left(\frac{W_c}{2.67 \times 10^{10}}\right)^{6/7} \tag{A7.5.10}$$

Putting the value of $N_D^{-3/4}$ from Eq. (A7.5.10) into Eq. (A7.5.8), we obtain

$$BV_{PP} = 5.34 \times 10^{13} \times \left(\frac{W_c}{2.67 \times 10^{10}}\right)^{6/7} = \frac{5.34 \times 10^{13} W_c^{6/7}}{(2.67 \times 10^{10})^{6/7}} \tag{A7.5.11}$$

Dividing the first term of eq. (A7.5.2), as given by Eq. (A7.5.7) by BV_{PP} given by Eq. (A7.5.11), we have

$$\begin{aligned} \text{First term}/BV_{PP} &= -\frac{(3.25 \times 10^{35})^{1/7}}{2} \times \frac{(2.67 \times 10^{10})^{6/7}}{5.34 \times 10^{13}} \times \left(\frac{r_j}{W_c}\right)^{6/7} \\ &= -0.9663806 \times \left(\frac{r_j}{W_c}\right)^{6/7} \cong -1 \times \left(\frac{r_j}{W_c}\right)^{6/7} = -\left(\frac{r_j}{W_c}\right)^{6/7} \end{aligned} \tag{A7.5.12}$$

Now, pre-logarithmic factor of second term of Eq. (A7.5.2) is

$$\text{Factor} = \frac{qN_D}{2\varepsilon_0 \varepsilon_s} r_d^2 \tag{A7.5.13}$$

Using Eq. (A7.5.3), we obtain

$$\frac{r_d^2 - r_j^2}{r_j} = \frac{2\varepsilon_0 \varepsilon_s E_{pCYL}}{qN_D} \quad \text{or} \quad \frac{r_d^2}{r_j} - r_j = \frac{2\varepsilon_0 \varepsilon_s E_{pCYL}}{qN_D}$$

$$\text{or} \quad \frac{r_d^2}{r_j} = r_j + \frac{2\varepsilon_0 \varepsilon_s E_{pCYL}}{qN_D} \quad \text{or} \quad \frac{r_d}{\sqrt{r_j}} = \sqrt{r_j + \frac{2\varepsilon_0 \varepsilon_s E_{pCYL}}{qN_D}}$$

$$\text{or}\quad r_{\mathrm{d}} = \sqrt{r_{\mathrm{j}}}\sqrt{r_{\mathrm{j}} + \frac{2\varepsilon_0\varepsilon_{\mathrm{s}}E_{\mathrm{pCYL}}}{qN_{\mathrm{D}}}} = \sqrt{r_{\mathrm{j}}}\times\sqrt{r_{\mathrm{j}}}\sqrt{1 + \frac{2\varepsilon_0\varepsilon_{\mathrm{s}}E_{\mathrm{pCYL}}}{qN_{\mathrm{D}}r_{\mathrm{j}}}}$$

$$= r_{\mathrm{j}}\sqrt{1 + \frac{2\varepsilon_0\varepsilon_{\mathrm{s}}E_{\mathrm{pCYL}}}{qN_{\mathrm{D}}r_{\mathrm{j}}}} \qquad \text{(A7.5.14)}$$

Substituting for r_{d} from Eq. (A7.5.14) into Eq. (7.5.13), we have

$$\text{Factor} = \frac{qN_{\mathrm{D}}}{2\varepsilon_0\varepsilon_{\mathrm{s}}}r_{\mathrm{j}}^2\left(1 + \frac{2\varepsilon_0\varepsilon_{\mathrm{s}}E_{\mathrm{pCYL}}}{qN_{\mathrm{D}}r_{\mathrm{j}}}\right) \qquad \text{(A7.5.15)}$$

Parallel-plane breakdown voltage is obtained from the potential distribution

$$V(x) = \frac{qN_{\mathrm{D}}}{2\varepsilon_0\varepsilon_{\mathrm{s}}}\left(2W_{\mathrm{c}}^2 - x^2\right) \qquad \text{(A7.5.16)}$$

by putting $x = W_{\mathrm{c}}$ = depletion region width at breakdown. Then Eq. (A7.5.16) reduces to

$$BV_{\mathrm{PP}} = \frac{qN_{\mathrm{D}}}{2\varepsilon_0\varepsilon_{\mathrm{s}}}\left(2W_{\mathrm{c}}^2 - W_{\mathrm{c}}^2\right) = \frac{qN_{\mathrm{D}}W_{\mathrm{c}}^2}{2\varepsilon_0\varepsilon_{\mathrm{s}}} \qquad \text{(A7.5.17)}$$

Dividing Eq. (A7.5.15) by Eq. (A7.5.17), we get

$$\text{Normalized prelogarithmic factor} = \text{Factor}/BV_{\mathrm{PP}} = \frac{r_{\mathrm{j}}^2}{W_{\mathrm{c}}^2}\left(1 + \frac{2\varepsilon_0\varepsilon_{\mathrm{s}}E_{\mathrm{pCYL}}}{qN_{\mathrm{D}}r_{\mathrm{j}}}\right)$$

$$= \frac{r_{\mathrm{j}}^2}{W_{\mathrm{c}}^2} + \frac{2\varepsilon_0\varepsilon_{\mathrm{s}}E_{\mathrm{pCYL}}}{qN_{\mathrm{D}}}\left(\frac{r_{\mathrm{j}}}{W_{\mathrm{c}}^2}\right) \qquad \text{(A7.5.18)}$$

Using Eq. (A7.5.10), we have

$$\frac{2\varepsilon_0\varepsilon_{\mathrm{s}}E_{\mathrm{pCYL}}}{qN_{\mathrm{D}}}\left(\frac{r_{\mathrm{j}}}{W_{\mathrm{c}}^2}\right) = \frac{2\varepsilon_0\varepsilon_{\mathrm{s}}}{q}\times\left(\frac{3.25\times10^{35}}{r_{\mathrm{j}}}\right)^{1/7}\times\frac{W_{\mathrm{c}}^{8/7}}{\left(2.67\times10^{10}\right)^{8/7}}$$

$$= \frac{2\times(8.854\times10^{-14})(11.9)}{1.6\times10^{-19}}\times\frac{\left(3.25\times10^{35}\right)^{0.143}}{\left(2.67\times10^{10}\right)^{1.143}}$$

$$\times r_{\mathrm{j}}^{1-1/7}\times W_{\mathrm{c}}^{8/7-2}$$

$$= 1.90674\left(\frac{r_{\mathrm{j}}}{W_{\mathrm{c}}}\right)^{6/7} \cong 2\left(\frac{r_{\mathrm{j}}}{W_{\mathrm{c}}}\right)^{6/7} \qquad \text{(A7.5.19)}$$

Applying Eq. (A7.5.14), the logarithmic term of eq. (A7.5.2) becomes

$$\ln\left(\frac{r_\mathrm{d}}{r_\mathrm{j}}\right) = \ln\left(\sqrt{1+\frac{2\varepsilon_0\varepsilon_\mathrm{s}E_\mathrm{pCYL}}{qN_\mathrm{D}r_\mathrm{j}}}\right) = \frac{1}{2}\ln\left(1+\frac{2\varepsilon_0\varepsilon_\mathrm{s}E_\mathrm{pCYL}}{qN_\mathrm{D}r_\mathrm{j}}\right) \quad \text{(A7.5.20)}$$

Substituting the values of E_pCYL and N_D from Eqs. (A7.5.6) and (A7.5.10), respectively, Eq. (A7.5.20) reduces to

$$\ln\left(\frac{r_\mathrm{d}}{r_\mathrm{j}}\right) = \frac{1}{2}\ln\left\{1+\frac{2\varepsilon_0\varepsilon_\mathrm{s}\left(\dfrac{3.25\times10^{35}}{r_\mathrm{j}}\right)^{1/7}}{q\left(\dfrac{2.67\times10^{10}}{W_\mathrm{c}}\right)^{8/7}r_\mathrm{j}}\right\}$$

$$= \frac{1}{2}\ln\left\{1+\frac{(2\times8.854\times10^{-14}\times11.9)\times\left(\dfrac{3.25\times10^{35}}{r_\mathrm{j}}\right)^{1/7}}{(1.6\times10^{-19})\times\left(\dfrac{2.67\times10^{10}}{W_\mathrm{c}}\right)^{8/7}r_\mathrm{j}}\right\}$$

$$\frac{1}{2}\ln\left\{1+\frac{(2\times8.854\times10^{-14}\times11.9)\times\left(\dfrac{3.25\times10^{35}}{r_\mathrm{j}}\right)^{1/7}}{(1.6\times10^{-19})\times(2.67\times10^{10})^{8/7}r_\mathrm{j}}\times\frac{W_\mathrm{c}^{8/7}}{r_\mathrm{j}^{1+1/7}}\right\}$$

$$= \frac{1}{2}\ln\left\{1+1.90675\left(\frac{W_\mathrm{c}}{r_\mathrm{j}}\right)^{8/7}\right\} \cong \frac{1}{2}\ln\left\{1+2\left(\frac{W_\mathrm{c}}{r_\mathrm{j}}\right)^{8/7}\right\} \quad \text{(A7.5. 21)}$$

Finally, combining together Eqs. (A7.5.12), (A7.5.18), (A7.5.19), and (A7.5.21), we get the desired equation for the ratio (breakdown voltage of cylindrical junction/breakdown voltage of parallel-plane junction) as

$$\frac{BV_\mathrm{CYL}}{BV_\mathrm{PP}} = \frac{1}{2}\left\{\left(\frac{r_\mathrm{j}}{W_\mathrm{c}}\right)^2+2\left(\frac{r_\mathrm{j}}{W_\mathrm{c}}\right)^{6/7}\right\}\ln\left\{1+2\left(\frac{W_\mathrm{c}}{r_\mathrm{j}}\right)^{8/7}\right\}-\left(\frac{r_\mathrm{j}}{W_\mathrm{c}}\right)^{6/7} \quad \text{(A7.5.22)}$$

REFERENCES

1. B. J. Baliga and S. K. Ghandhi, Analytical Solutions for the Breakdown Voltage of Abrupt Cylindrical and Spherical Junctions, *Solid State Electron.*, Vol. 19, No. 9, September 1976, pp. 739–744.
2. B. J. Baliga, *Power Semiconductor Devices*, PWS Publishing Co., Boston, 1996.

8

IGBT PROCESS DESIGN AND FABRICATION TECHNOLOGY

IGBTs are very large scale integrated circuits consisting of millions of microdevices working in parallel. Techniques like electron beam writing on wafer are used for the close tolerances and high resolution needed for IGBT processing. Microelectronics is the technological vehicle of IGBT fabrication. At the same time, techniques like surface field control, and so on, from the realm of large-area power device fabrication are also employed for IGBT realization. Manufacturing process of an IGBT fabrication line involves not only IC and power discrete technologies but also MOS and bipolar process steps together, [1–7]. The boundaries of these disciplines are disappearing with their merger to produce combination devices.

It is imperative for high-performance IGBT fabrication to keep MOS channel lengths short (≤ 1 μm). The combined formation of a submicrometer channel MOS device, integrated with a thyristor structure, requires trade-off analyses of several complex process parameters. The use of results from one-dimensional and multidimensional numerical process simulation studies, is a helpful guide toward efficient and successful process design. To fulfill the combined requirements of MOSFET and bipolar device processing, high quality, low Q_{ss} (Section 7.1.1), MOS oxide, must be grown and preserved throughout the IGBT fabrication process, while at the same time,

---

The Insulated Gate Bipolar Transistor (IGBT) Theory and Design, By Vinod Kumar Khanna.
ISBN 0-471-23845-7 © 2003 Institute of Electrical and Electronics Engineers

stringent control of minority-carrier lifetime must be maintained, to preserve it at a high value, to ensure the conductivity modulation of the thick N^--base region.

8.1 PROCESS SEQUENCE DEFINITION

8.1.1 VDMOSFET IGBT Fabrication

The device structural design and process design are carried out side-by-side, with due consideration of technological constraints and including only feasible solutions. Figure 8.1 presents the main steps in the fabrication process sequence of a punchthrough (PT)-IGBT (Section 2.1.2). For easy visualization, the isometric diagrams at each stage are presented in Fig. 8.2. The starting material is high-purity, dislocation-free Czochralski (CZ)/float zone (FZ) silicon (Section 7.1.1), having orientation ⟨100⟩ and heavily doped with P-type impurity to form a P^+ substrate, step (a). (For NPT-IGBT, neutron-transmutation doped (NTD) N-type float zone Si provides tight tolerances for the dopant density fluctuations required to achieve high manufacturing yields for large-area IGBT chips. In the finished devices, this low-doped substrate constitutes the N^--base layer, blocking the high voltage, when the device is in the OFF state). After cleaning and polishing of the P^+ substrate, two N-type layers are grown by epitaxy [step (b)]. The first layer is the N-buffer layer. In the second layer, the device will be formed. The dopant density and thickness of this epitaxial layer have been pre-decided by the voltage rating. As the voltage rating of the IGBT is raised, the thickness and resistivity of the N^- epilayer increase.

For the first photolithographic step, step (d), the surface is thermally oxidized (oxide thickness $\sim 1\ \mu$m), step (c). This defines the P^+ region in the center of P base. When the deep P^+ region is produced, a similar P^+ region is created for the field ring for high-voltage termination structure around the periphery of the chip. The thick oxide grown here is called the *field oxide* and prevents any inversion of the underlying Si on applying any voltage. After etching windows through the oxide, thin scattering oxide is grown ~ 500 Å, step (e). The scattering oxide serves a dual purpose. Ions can penetrate deep into crystalline silicon if the ion beam strikes the lattice in a direction, which allows them to channel along planes containing a small number of atoms. Such channeling is undesirable because it is sensitive to the angle of incidence of the ion beam. Coating the silicon surface with the thin scattering oxide layer helps in preventing channeling due to the amorphous nature of SiO_2. Besides scattering the incident ion beam to provide uniform deposition, it helps in shifting the peak impurity concentration at the surface of the wafer. This is because in ion implantation, the peak impurity concentration lays a few hundred angstroms below the wafer surface. As the scattering oxide is removed after implantation, this peak, initially situated below the surface, will be subsequently relocated at the surface. Boron implantation, step (f), is done at a low energy 60 keV and high dose (10^{16} cm^{-2}) to furnish a pre-deposited layer very rich in boron (dose = number of dopant atoms per cm^2). Then in the drive-in to be done afterwards, the boron content will not be much depleted and the layer will remain conducting. This is necessary because this step is introduced to reduce the resistance of the P-base region to avert latchup of the IGBT. After annealing and scattering oxide etching, step (g), the implanted boron is driven-in, step (h), to achieve a depth of 6 μm (suppose) in final device. It may be noted here that the edges of the field ring shown in mask I, step (d), are rounded (not straight, as shown) because sharp corners lead to electric field crowding and hence premature breakdown.

This is followed by a second photolithography for defining the active area, step (i). "Active area" means the pattern of cells, square or hexagonal. These cells are centered on the previously formed P^+ regions. In the drawings, only one cell is shown. In an actual device, there will be $\sim 10^5$ or more cells per sq.cm. The next two stages are gate oxidation and polysilicon deposition, step (j). The gate oxide is 500–1000 Å thick and is grown in clean chlorine ambient—for example, dry O_2 using trichloroethylene as the chlorine source. This is essential because a high surface-state density may raise threshold voltage to an undesirable magnitude.

Then third photolithography, step (k), defines the P bases of the cells. After etching and screen oxidation, step (l), boron implantation, step (m), is done at a higher energy ($\sim$ 80 keV) and lower dose (10^{14} cm^{-2}). Then the wafers are annealed in nitrogen and screen oxide is etched away, step (n). Boron is driven-in, step (o), to a depth of 4 μm (for example) in the final device.

Thereafter, the fourth photolithography step, step (p), is carried out for phosphorus diffusion in emitter regions. As before, following screen oxidation, step (q), phosphorus or arsenic implantation, step (r), is done at 50 keV, 10^{15} cm^{-2}. The succeeding steps are annealing, screen oxide removal, step (s), and the phosphorus drive-in, step (t), to a depth of 1 μm.

Two PN junctions (N^+ emitter/P base and P base/N^- base) define the IGBT channel. Each junction follows the edge of the gate mask. Because both the ends of the channel are defined with the help of the gate structure as a mask, accurate alignment of gate and channel is ensured. Also, channel length is controlled with great precision. This is the advantage of the self-aligned process.

One-micron-thick CVD SiO_2, doped heavily $\sim$2–10% by weight with phosphorus (hence known as phosphosilicate glass or PSG), is deposited on the wafers. It serves as an intermediate dielectric providing insulation between the polysilicon and the metal. The phosphorus present inside the glass decreases its viscosity. So, the glass flows easily at relatively low temperatures (950–1050°C). This improves the step coverage by the metal layer by reducing the sharpness of the contact hole walls and the steps that occur at the

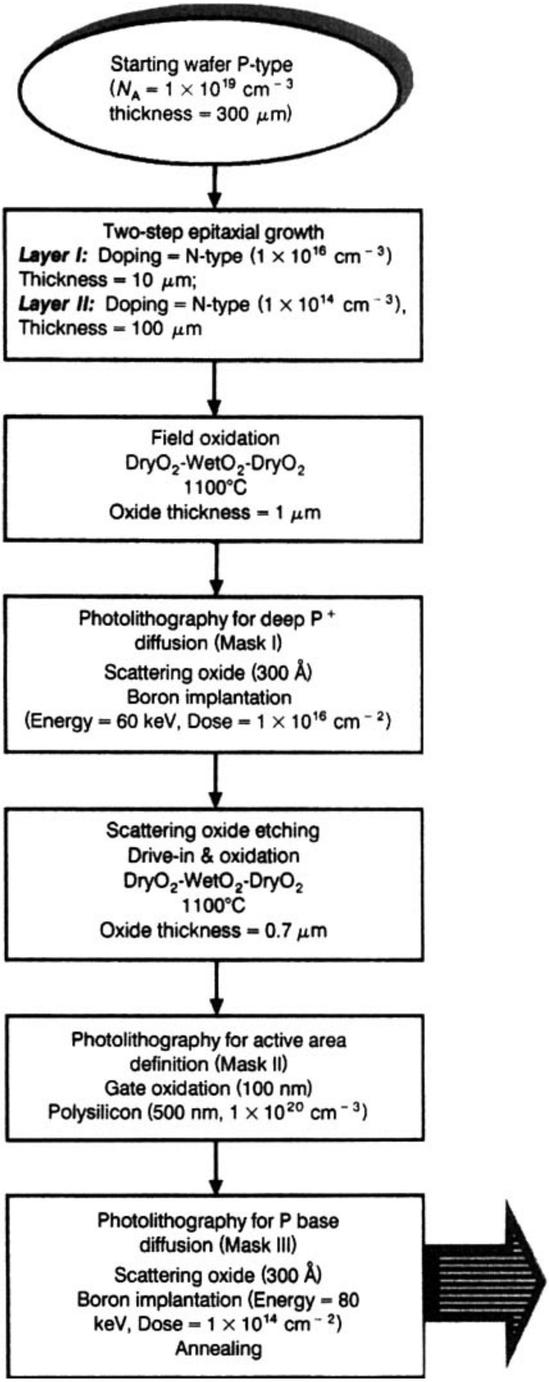

Starting wafer P-type
($N_A = 1 \times 10^{19}$ cm^{-3}
thickness = 300 μm)
Two-step epitaxial growth
Layer I: Doping = N-type (1×10^{16} cm^{-3})
Thickness = 10 μm;
Layer II: Doping = N-type (1×10^{14} cm^{-3}),
Thickness = 100 μm
Field oxidation
DryO$_2$-WetO$_2$-DryO$_2$
1100°C
Oxide thickness = 1 μm
Photolithography for deep P$^+$ diffusion (Mask I)
Scattering oxide (300 Å)
Boron implantation
(Energy = 60 keV, Dose = 1×10^{16} cm^{-2})
Scattering oxide etching
Drive-in & oxidation
DryO$_2$-WetO$_2$-DryO$_2$
1100°C
Oxide thickness = 0.7 μm
Photolithography for active area definition (Mask II)
Gate oxidation (100 nm)
Polysilicon (500 nm, 1×10^{20} cm^{-3})
Photolithography for P base diffusion (Mask III)
Scattering oxide (300 Å)
Boron implantation (Energy = 80 keV, Dose = 1×10^{14} cm^{-2})
Annealing

Figure 8.1 Flow chart of (VDMOS-PT) vertical double-diffused MOS-punchthrough IGBT fabrication. Dry O_2–WetO_2–Dry O_2 thermal oxidation cycle involves growing oxide by dry oxygen; then by bubbling oxygen through water at 95°C or reacting hydrogen and oxygen inside the furnace tube to form water (pyrogenic oxidation); and finally again by dry oxygen.

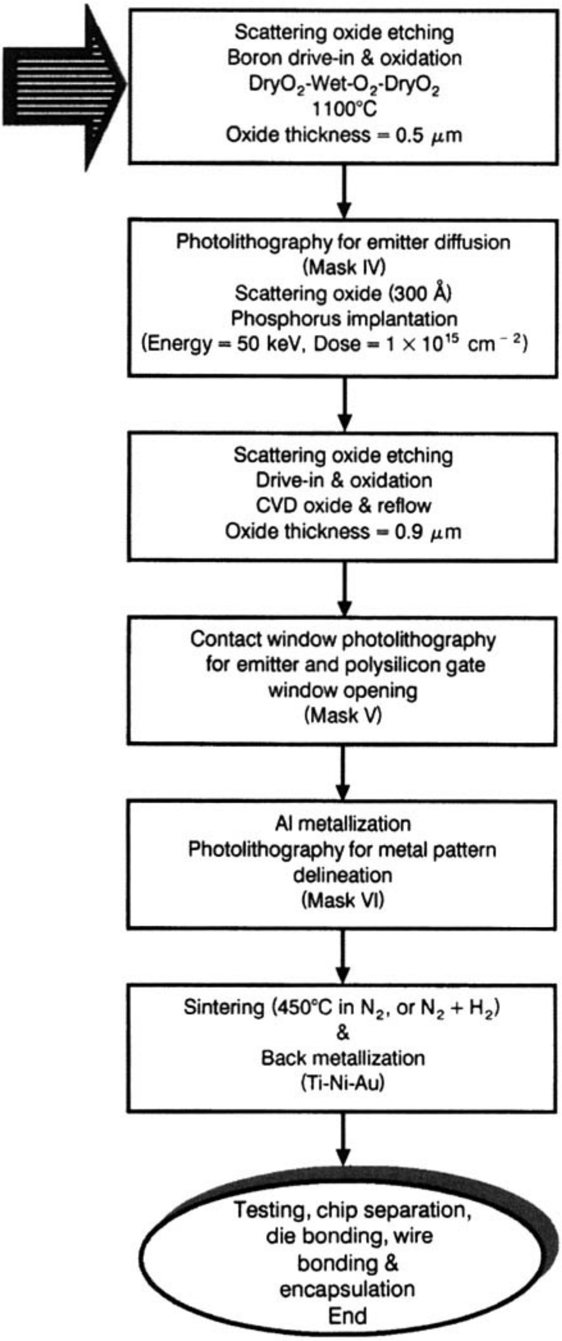
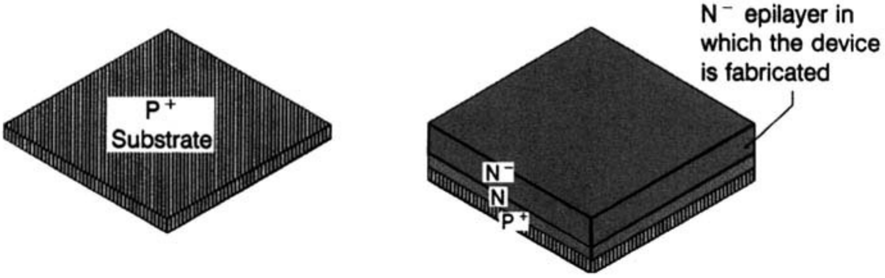
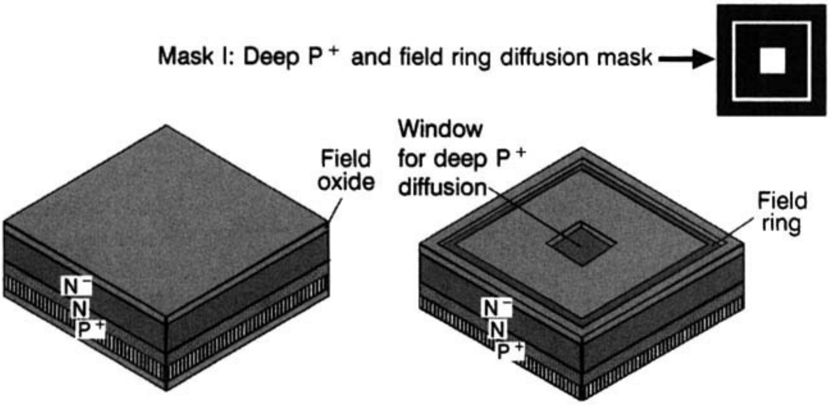

Scattering oxide etching
Boron drive-in & oxidation
$DryO_2$-Wet-O_2-$DryO_2$
1100°C
Oxide thickness = 0.5 μm
Photolithography for emitter diffusion
(Mask IV)
Scattering oxide (300 Å)
Phosphorus implantation
(Energy = 50 keV, Dose = 1×10^{15} cm^{-2})
Scattering oxide etching
Drive-in & oxidation
CVD oxide & reflow
Oxide thickness = 0.9 μm
Contact window photolithography
for emitter and polysilicon gate
window opening
(Mask V)
Al metallization
Photolithography for metal pattern
delineation
(Mask VI)
Sintering (450°C in N_2, or $N_2 + H_2$)
&
Back metallization
(Ti-Ni-Au)
Testing, chip separation,
die bonding, wire
bonding &
encapsulation
End

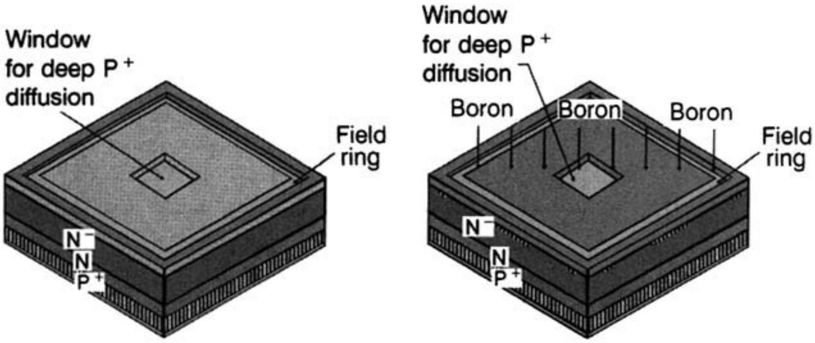
Figure 8.1 (*Continued*).


P+
Substrate
N⁻ epilayer in which the device is fabricated
N⁻
N
P+

(a) Starting P$^+$ substrate wafer (b) Growth of N and N$^-$ epilayers

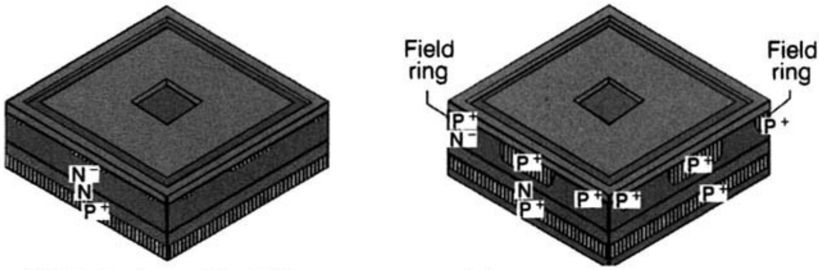

Mask I: Deep P+ and field ring diffusion mask
Field oxide
N⁻
N
P+
Window for deep P+ diffusion
Field ring
N⁻
N
P+

(c) Field oxidation (d) Photolithography for deep P$^+$ diffusion and field ring (mask I)


Window for deep P+ diffusion
Field ring
N⁻
N
P+
Window for deep P+ diffusion
Boron
Boron
Boron
Field ring
N⁻
N
P+

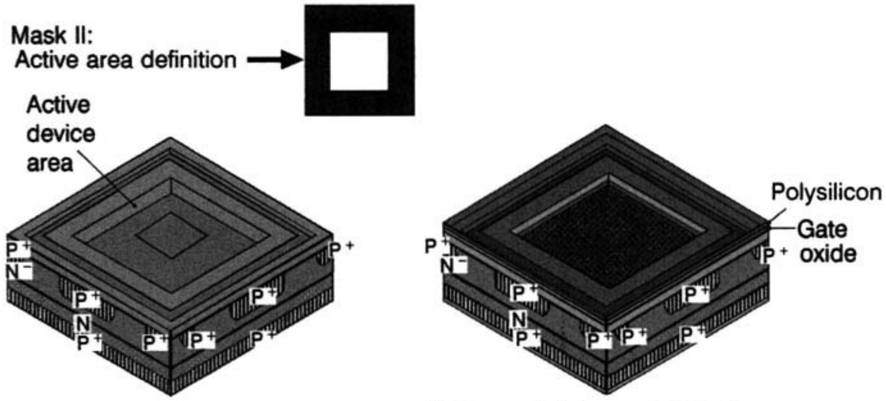
(e) Scattering oxide growth (f) Boron implantation

Figure 8.2 Cross-sectional view of the structures during different stages of VDMOS PT-IGBT fabrication. Unit cell with field ring is shown; plan view of all masks is also given.


N$^-$
N
P$^+$
Field ring
Field ring
P$^+$
N$^-$
P$^+$
P$^+$
N
P$^+$
P$^+$
P$^+$
P$^+$
P$^+$

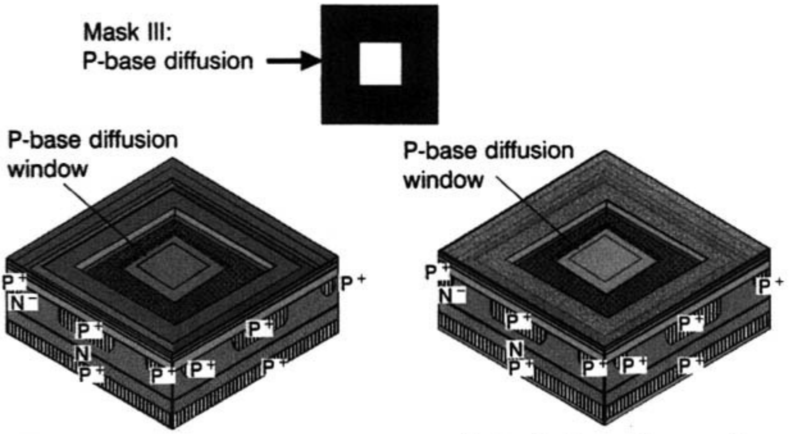
(g) Scattering oxide etching

(h) Drive-in and oxidation


Mask II:
Active area definition
Active device area
P$^+$
N$^-$
P$^+$
P$^+$
P$^+$
N
P$^+$
P$^+$
P$^+$
P$^+$
Polysilicon
Gate oxide
P$^+$
N$^-$
P$^+$
P$^+$
N
P$^+$
P$^+$
P$^+$
P$^+$
P$^+$

(i) Photolithography for active area definition (mask II)

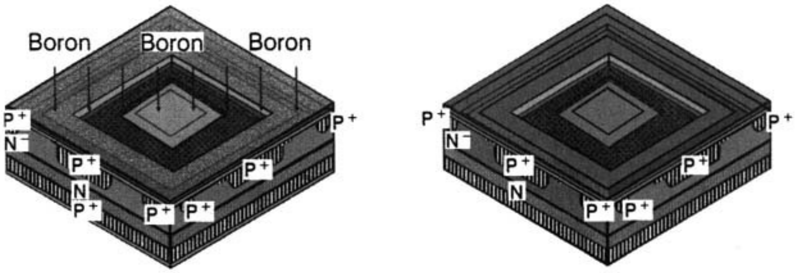
(j) Gate oxidation and Polysilicon deposition


Mask III:
P-base diffusion
P-base diffusion window
P-base diffusion window
P$^+$
N$^-$
P$^+$
P$^+$
P$^+$
N
P$^+$
P$^+$
P$^+$
P$^+$
P$^+$
N$^-$
P$^+$
P$^+$
P$^+$
N
P$^+$
P$^+$
P$^+$
P$^+$

(k) Photolithography for P-base diffusion (mask III)

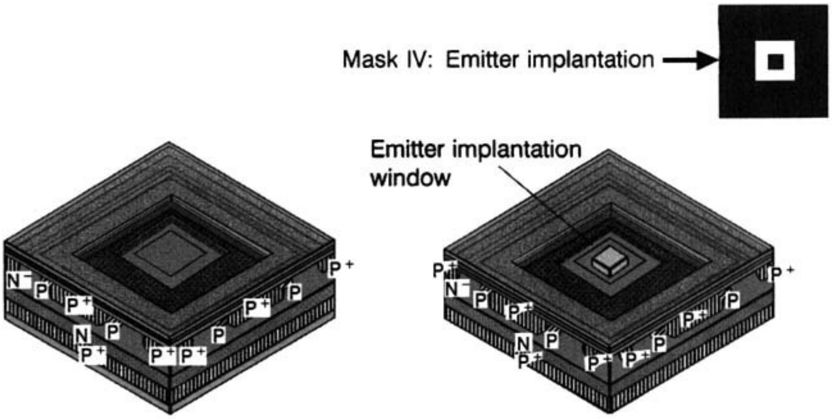
(l) Scattering oxide growth

Figure 8.2 (*Continued*).


Boron
Boron
Boron
P+
N−
P+
P+
N
P+
P+
P+
P+
P+
N−
P+
P+
N
P+
P+
P+

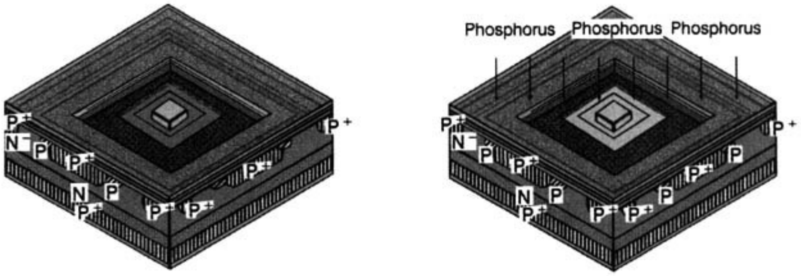
(m) Boron implantation

(n) Scattering oxide etching


Mask IV: Emitter implantation
Emitter implantation window
N−
P
P+
P+
P
N
P
P+
P
P+
P+
P+
P+
N−
P
P+
P+
P
N
P
P+
P
P+
P+
P+

(o) Drive-in and oxidation

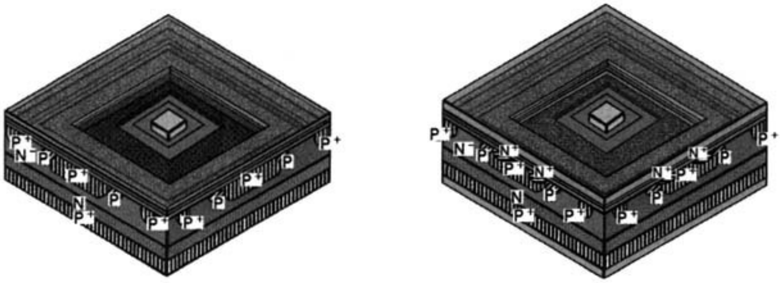
(p) Photolithography for emitter implantation (mask IV)


P+
N−
P
P+
P+
N
P
P+
P+
P+
P+
Phosphorus
Phosphorus
Phosphorus
P+
N−
P
P+
P+
P
N
P
P+
P
P+
P+
P+

(q) Scattering oxide growth

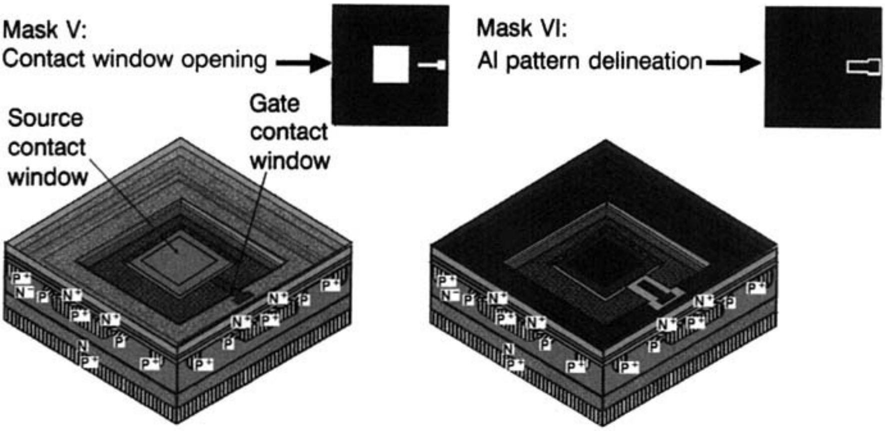
(r) Phosphorus implantation

Figure 8.2 (*Continued*).



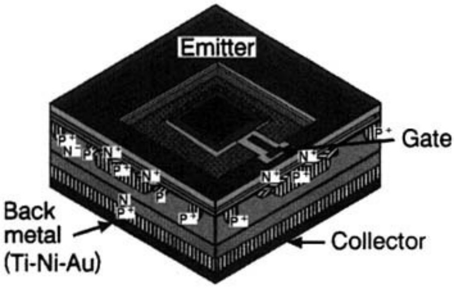
(s) Scattering oxide etching

(t) Drive-in and oxidation


Mask V:
Contact window opening
Source contact window
Gate contact window
Mask VI:
Al pattern delineation
P+
N−
P
N+
N

(u) Photolithography for contact window opening (mask V)

(v) Front Al metallization and patterning (mask VI)


Emitter
Gate
Back metal (Ti-Ni-Au)
Collector
P+
N−
P
N+
N

(w) Back metallization

Figure 8.2 (*Continued*).

polysilicon edges. This oxide can be easily made to flow and form smooth glassy contours over the edges of polySi. The conformal coating of oxide helps in smoothening the surface topography, and it allows the continuity of metallization stages to follow ahead in the process. Inclusion of phosphorus also protects the device from mobile Na^+ and other ionic contaminants. This step is not shown in Fig. 8.2.

The fifth photolithographic step, step (u), opens the contact windows to establish contact with the source as well as with the polysilicon gate layer. The window is etched down to the polysilicon layer over the gate so that the aluminum electrode can make contact there as well.

Aluminum is deposited by vacuum evaporation or sputtering, step (v). The sixth photolithography step, step (v), is used for Al pattern delineation. Al sintering in nitrogen or forming gas alloys the metal with Si. Hydrogen present in the forming gas aids in annealing out any radiation damage produced during metal deposition. On the backside of the wafer, preferably a solderable multilayer metallization scheme like Ti–Ni–Au, Cr–Ni–Ag, or Cr–Au is deposited, step (w). In these metallization schemes, the bottom layer such as titanium or chromium (thickness = 3000 Å) serves the purpose of a barrier layer, preventing the metal to diffuse into silicon, and also eats away any remnant native oxide on silicon, improving film adhesion. The second layer such as Ni (5000 Å) provides solderable contact. The flash of Au/Ag (2000 Å) gives uniform wetting of the surface and prevents the metallization from environmental degradation. Gold itself is solderable; and if Ni layer is absent, the gold layer must be thicker for good contact. The final capping layer of the device is a passivating silicon nitride layer deposited by plasma-enhanced CVD (PECVD). It seals and thereby protects the wafer surface from mechanical scratches or the influence of contaminants. Then windows are etched in this capping layer for making external wire connections. The wafer is electrically tested and diced for chip separation. The chips are mounted on the package with the help of solder preforms, and electrical contacts on the top surface are made by wire bonding thick Au wires according to current rating.

On careful inspection of the process, it is noticed that there are three drive-in and oxidation steps: (h) for deep P^+, (o) for P base, and (t) for N^+ emitter. Their temperatures and times are adjusted according to the depth of the P^+, P, or N^+ layer and the oxide thickness needed for the succeeding process step requirement. The process and resulting oxide parameters are given in Table 8.3. Similarly, there are two boron implantation steps, (m) and (f), which must be differentiated. In step (m), deep P^+ layer is produced in the center of the cell, while in step (f), a shallower P layer is created for making the P base of the cell. Naturally, therefore, the implantation energy and dose are different for these steps (Fig. 8.1 and Table 8.3).

Unlike a power MOSFET, the switching speed of an IGBT must be enhanced for catering to the high-frequency applications. Electron irradiation technique (at 3 MeV energy and fluence from 0.5 to 16 Mrad) has been used to control the switching speed over broad latitude of turn-off times from 20 μsec to 200 nsec (*Fluence* of atomic or nuclear particles is defined as the time integral of their flux density. Hence, it measures the time-integrated particle flux. Its unit is the number of particles per square centimeter).

As an increase of forward drop and leakage current of devices invariably accompany the introduction of deep levels, efforts are made to seek a trade-off amongst these parameters, best suited for a given application. Investigations have shown that as the turn-off time decreases, the forward drop slowly increases in the beginning until a critical point is reached, after which any further reduction of turn-off time causes a steep rise of forward drop. The influence of a given radiation dose on turn-off time is more pronounced for the higher breakdown voltage devices owing to their lighter doping and larger N^--base thickness. The effect of radiation in the above dose levels on leakage current was nominal. It remained in the nanoampere range after irradiation so that the blocking performance of the device was not impaired.

Critical Processing Steps

(i) In the photolithographic step for phosphorus implantation to produce the N^+ emitter, (Fig. 8.2p), misalignment of the oxide island with respect to the edges of the polysilicon may increase the length of the N^+ emitter, resulting in degradation of dV/dt and safe operating area.

(ii) In the contact window opening photolithography (Fig. 8.2u), misalignment either leads to overlapping of the contact metal with polysilicon or shortening of the metal layer over the P-base region. In the former case, the emitter is shorted with the gate, while in the latter the contact with P base is poor.

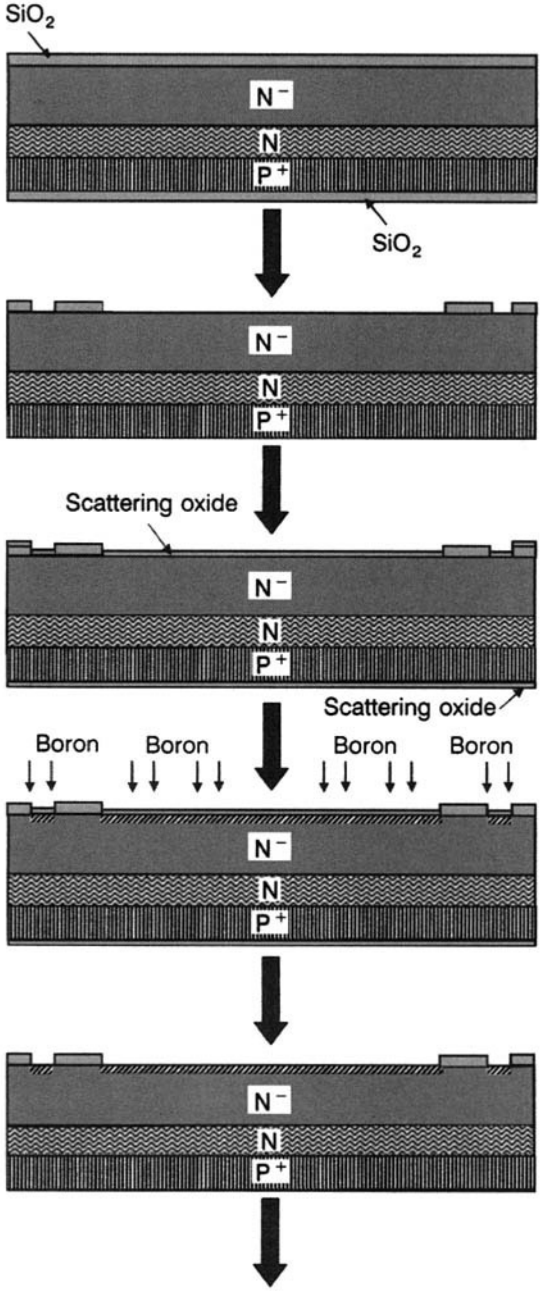
Localized lifetime tailoring of the interface between N-drift layer and N-buffer layer allows the fabrication of more reliable PT-IGBTs. Recently, the development of a new punchthrough IGBT has been reported [8]. This IGBT gives more rugged SOA and improved short-circuit endurance like NPT structure. It also exhibits superior trade-off relation between ON-state voltage and turn-off loss than NPT-IGBT. But the new IGBT has a longer tail time due to its thicker N^--drift layer. Nevertheless, its turn-off loss is smaller because its tail current is lower and also its length is shorter.

For fabrication of this IGBT structure, the starting wafer is an N-type silicon wafer. A 100-μm-thick P-collector layer is formed by diffusion. The thickness of the N-type wafer is taken such that at the specified voltage the depletion layer touches the P collector. This thin N^--drift layer minimizes the saturation voltage. But its thickness is about 10 μm more than that of conventional PT-IGBT to improve the destruction tolerance. The N-buffer layer is formed by proton irradiation near the interface between P collector and N^- layer, followed by annealing. By converting the proton-irradiated area into a donor region, the buffer layer with lower resistivity than the N^- layer is produced.

Thus the fabrication of this IGBT utilizes localized lifetime killing by proton irradiation (Section 8.2.10) in place of reduction of lifetime of the whole wafer by electron beam irradiation (Section 8.2.9).

8.1.2 Trench–Gate IGBT Fabrication

The process for the trench–gate IGBT [9] diverges from the step of growing the field oxide, followed by patterning, to open an active area for boron implantation (Fig. 8.3). Here deep P^+ diffusion is not done. So, the first masking step produces the active area of the device along with the field ring(s). Main steps in the fabrication of trench–gate IGBT include: P-base and field–ring diffusion, N^+-emitter diffusion, trench formation, poly refilling into the trench, planarization, and metallization. After growing the field oxide over the whole wafer surface, step (a), the first photolithographic step, step (b), is performed for making windows through which P base as well as the peripheral field ring structure are implanted. This is accomplished using Mask I labeled as the P base and field ring implantation mask. Then the thin scattering oxide is grown over the entire wafer surface, step (c). Boron implantation is carried out, step (d). The energy and dose parameters for this boron implant are the same as for the second boron implant, step (m) in Fig. 8.2, for VDMOS-IGBT fabrication. After annealing, the scattering oxide is removed, step (e), and boron is driven in to achieve the P-base diffusion depth, step (f). The second photolithographic step, step (g), is intended for opening the regions where the N^+ emitter will be made by ion implantation. This is done using Mask II, which is called the N^+ emitter implantation mask. The scattering oxide is regrown, step (h) and phosphorus implantation is done, step (i). The process parameters for this step are the same as for step (r) in Fig. 8.2. Annealing follows the ion implantation step. Next, the scattering oxide is etched, step (j), and the N^+ impurity is driven in to obtain the designed N^+ emitter depth $= 1\ \mu\text{m}$, step (k). After removing the oxide from both sides of the wafer, silicon nitride film is formed by low-pressure chemical vapor deposition (LPCVD) followed by low-temperature CVD (LTCVD) oxide film for protection of the silicon wafer surface over P-base regions during the trench etching, which will be carried out after diffusion of N^+ emitter, step (l). Then comes the trench etching step. This involves the third photolithographic step, step (m). Using Mask III, the trench definition mask, the trench areas are delineated. Anisotropic etching by reactive ion etching (RIE) machine is used to excavate 5-μm-deep rectangular trenches, step (n). Trench depths of ~ 4–$10\ \mu\text{m}$ do not affect the breakdown voltage. Since in RIE the precision of trench depth can be ensured to within 10–20% of the designed target, the noncritical value of channel depth makes device


SiO$_2$
N$^-$
N
P$^+$
SiO$_2$
Scattering oxide
Scattering oxide
Boron
Boron
Boron
Boron

(a) Structure after step (c), field oxidation in Fig. 8.2

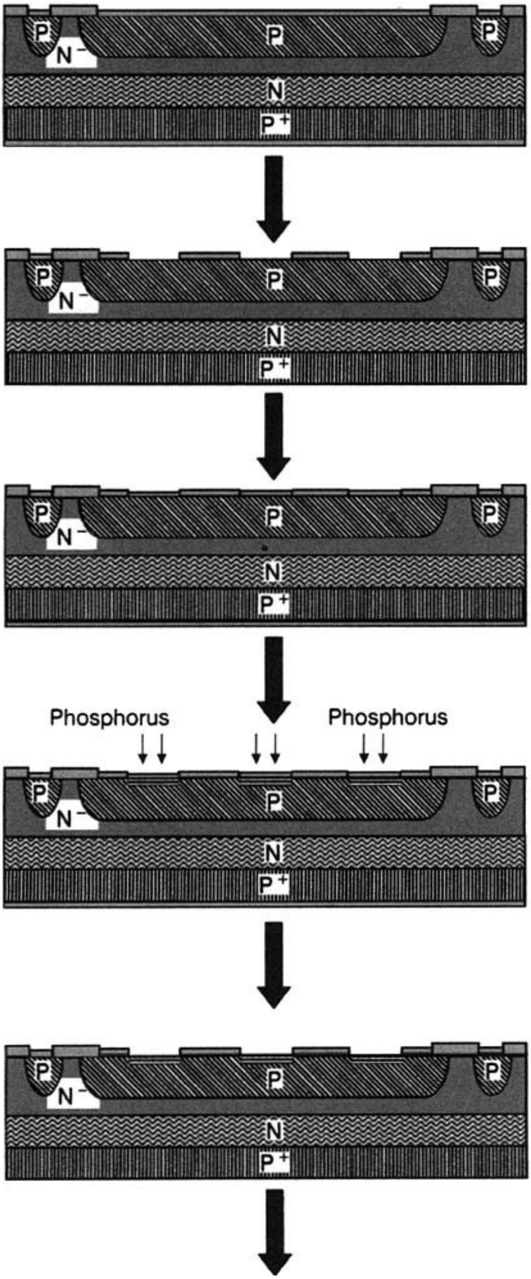
(b) Photolithography for an active area definition and field ring termination (mask I: P base and field ring implantation mask)

(c) Scattering oxide growth

(d) Boron implantation

(e) Scattering oxide etching

Figure 8.3 Representation of the device structure at various stages of trench-gate IGBT fabrication.


P
N⁻
P
P
N
P⁺
Phosphorus
Phosphorus

(f) Drive-in and oxidation

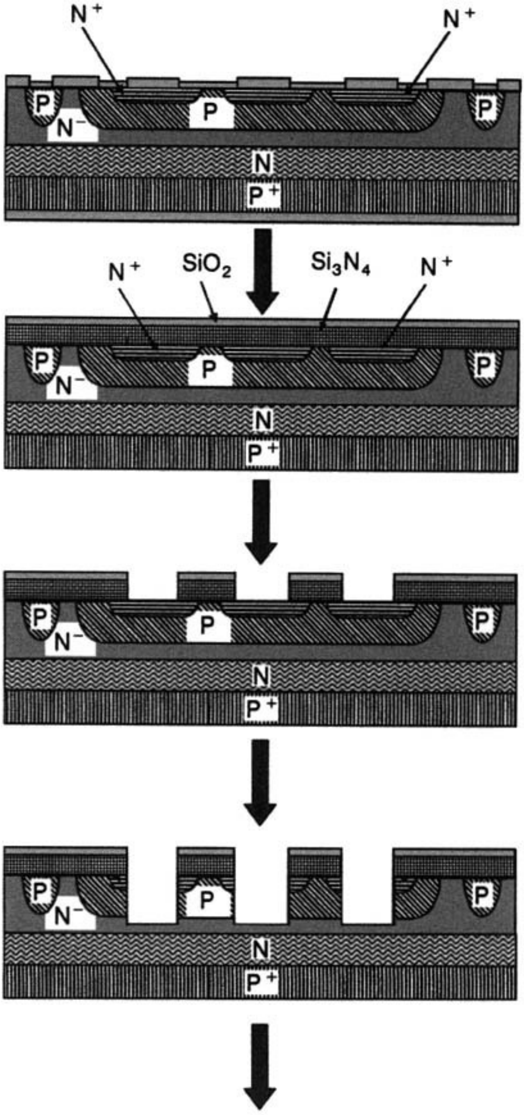
(g) Photolithography for emitter diffusion (mask II: N$^+$ emitter implantation mask)

(h) Scattering oxide growth

(i) Phosphorus implantation

(j) Scattering oxide etching

Figure 8.3 (*Continued*).



(k) Drive-in and oxidation

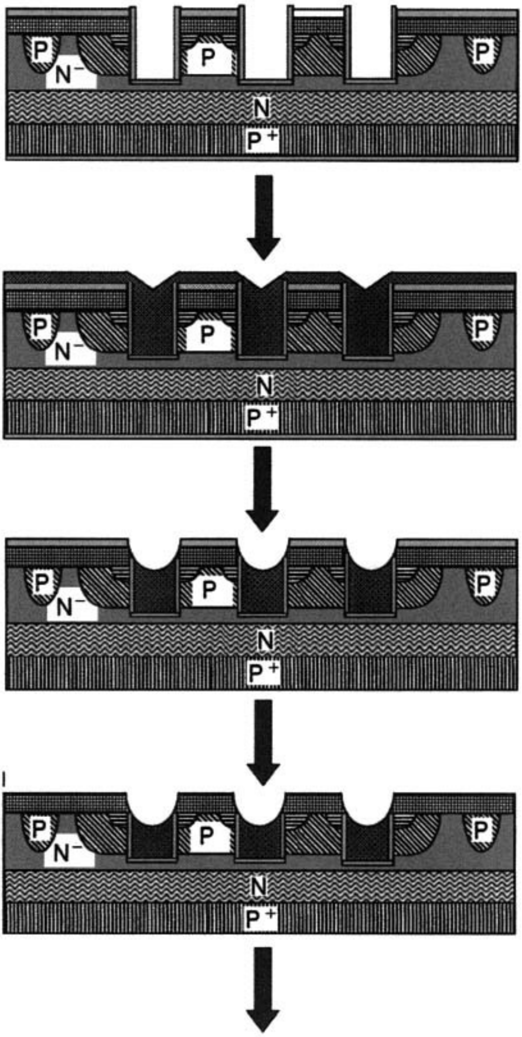
(l) Oxide etching from both sides, Nitride and oxide deposition

(m) Photolithography for trenching (mask III)

(n) Trench formation

Figure 8.3 (*Continued*).

fabrication easy. Although the trench width should desirably be minimized, thereby maximizing cell density, the minimum trench width is defined by process constraints and is typically 3 μm. During trench formation process, the surface of the silicon on the trench walls is subjected to the detrimental chemical effects of the reactant ions. It also faces their mechanical impact. To remove these damages, a sacrificial oxide layer about 1000 Å thick is grown by thermal oxidation. This layer is etched off and the damaged surface



(o) Gate oxidation

(p) Poly filling

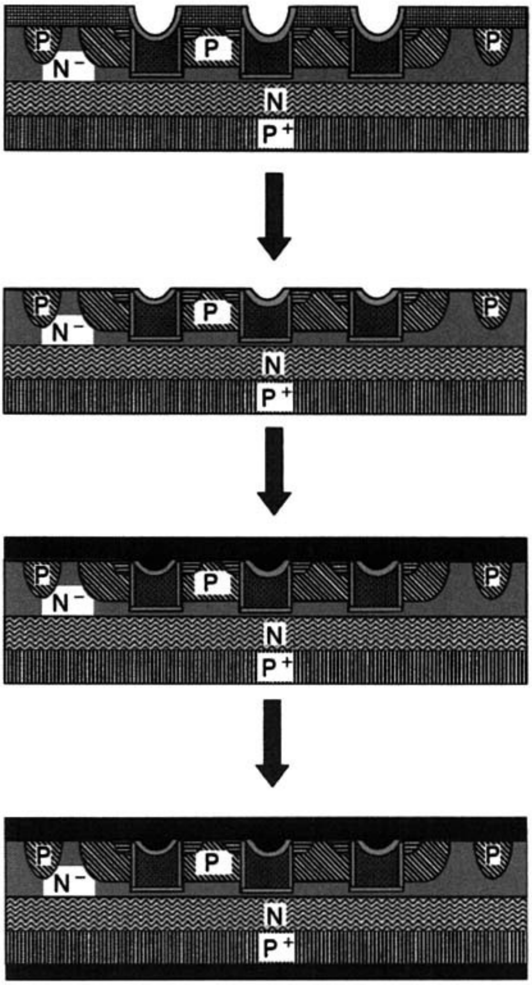
(q) Photolithography for polysilicon etching (mask IV) and planarization

(r) Oxide etching

Figure 8.3 (*Continued*).

layer is thereby eliminated. After the damages have been removed, the gate oxide is grown on the walls of the trench, step (o). This constitutes the gate oxidation step. Then trench is refilled by redepositing polysilicon into the trench, step (p). PolySi layer ~ 2 μm thick is deposited for trench refilling.

Because the surface of the silicon wafer at this stage is not plain but has high and low regions, a planarization process is carried out by reactive ion etching to etch back the polySi. Using Mask IV, a polysilicon etching mask, polysilicon is removed from everywhere except the gate regions, step (q). This is meant to etch back the polysilicon to form the gate. In step (r), the oxide is etched away. The next step is selective oxidation of polySi layer due to the



(s) Selective oxidation at gate regions

(t) Nitride etching

(u) Contact window opening photolithography (mask V) front metallization, metal patterning photolithography (mask VI), and Al sintering

(v) Back metallization

Figure 8.3 (*Continued*).

presence of nitride film over the mesa region, step (s). Now, silicon nitride film covers the mesa regions of the structure. So, upon oxidation, the polysilicon layer over the gate is selectively oxidized, leaving the silicon nitride film as such, which is then etched away by reactive ion etching, step (t). As in Fig. 8.2, Mask V and VI are respectively the contact window opening and metal pattern delineation masks. After fifth photolithographic step (u) using Mask V, aluminum is sputtered. Then in the sixth photolithographic step, step (u), using Mask VI, the aluminum is etched to make electrodes for emitter and gate of the IGBT. After sintering the aluminum to make good contact, the backside trilayer metal (Cr–Ni–Ag) is deposited, step (v). This completes the trench–gate fabrication process, to be followed by dicing, testing, and packaging.

8.2 UNIT PROCESS STEPS

8.2.1. Epitaxial Deposition

Epitaxial deposition is the deposition of a layer of single-crystal silicon on a substrate, such that the crystal structure of the deposited layer is an extension of the substrate [10, 11]. The Greek meaning of "epitaxy" is "upon ordered." Epitaxial deposition is a prime process step in the realization of punchthrough-IGBT (Section 2.1.2, Figs. 8.1–8.3). It is used for the growth of N^- base and N-buffer layers of the IGBT. It is this N^- base in which the P base and N^+ emitter of the IGBT are made. The critical IGBT structure therefore resides in N^- base. This requires that the epitaxial deposition process should be of high quality, resulting in the production of defect-free layers.

For Si, vapor-phase epitaxy is the most widely accepted. For growing silicon by epitaxy, four silicon sources have been used. These are silicon tetrachloride ($SiCl_4$), dichlorosilane (SiH_2Cl_2), trichlorosilane ($SiHCl_3$), and silane (SiH_4). Amongst these, $SiCl_4$ has found the widest industrial application. Epitaxial silicon is deposited by the hydrogen reduction of silicon tetrachloride gas at 1150–1300°C:

$$SiCl_4(Gas) + 2H_2 \longrightarrow Si(Solid) + 4HCl(Gas)$$

The overall reaction involves the following subreactions, with the species $SiHCl_3$ and SiH_2Cl_2 acting as intermediates:

$$\begin{aligned} &SiCl_4 + H_2 \rightleftharpoons SiHCl_3 + HCl, SiHCl_3 + H_2 \\ &\qquad \rightleftharpoons SiH_2Cl_2 + HCl, SiH_2Cl_2 \rightleftharpoons SiCl_2 + H_2 \\ &SiHCl_3 \rightleftharpoons SiCl_2 + HCl, SiCl_2 + H_2 \rightleftharpoons Si + 2HCl \end{aligned}$$

All these reactions are reversible, so that under appropriate conditions the corresponding backward reactions compete with the forward reactions, causing etching of silicon and removing the silicon from the substrate. From the temperature dependence of silicon growth rate, it is found that at lower temperatures (900–1100°C), the silicon growth is limited by reaction rate, while at higher temperatures (> 1100°C) the silicon growth is limited by transport processes, either by the amount of reactant reaching the silicon surface or by the products diffusing away. In this regime, known as the *mass transport* or *diffusion-limited region*, the growth rate is directly proportional to the partial pressure of the reactant in the carrier gas. Industrial epitaxial process is performed in this regime for minimal influence of temperature variations on the growth rate.

Epitaxial deposition is described by two processes, namely, *nucleation and growth*. Nucleation involves the creation of a small region of a new phase on another phase. The epitaxial process progresses by impingement, adsorption, diffusion, reaction, and desorption of the reacting species. Dopants are introduced into the epitaxial layer either simultaneously or intermittently through hydrides such B_2H_6, PH_3, AsH_3, and so on, diluted with a large quantity of hydrogen. The epitaxial deposition equipment comprises a gas distribution system, a high-purity quartz reactor tube, systems for substrate heating, electrical control, cooling, and exhaust.

8.2.2 Thermal Oxidation

In the IGBT process, thermal oxidation is used as a masking layer for impurity implantation/diffusion. It is also used as the gate dielectric. The field oxide in the IGBT prevents the action of any spurious voltages in inverting the underlying silicon. Besides, silicon dioxide forms a natural passivating cover over the planar N^+-emitter/P-base and P-base/N^--base junctions.

Thermal oxidation is done in an atmosphere containing oxygen or water vapor at high temperatures (900–1200°C):

$$\mathrm{Si} + \mathrm{O}_2 = \mathrm{SiO}_2$$

or

$$\mathrm{Si} + 2\mathrm{H}_2\mathrm{O} = \mathrm{SiO}_2 + 2\mathrm{H}_2$$

During this process, the oxidizing species (O_2 or H_2O) diffuses through the existing oxide layer, [12–15]. To produce an oxide layer of thickness d, a silicon layer of thickness 0.44 d is consumed. For the same temperature and time, oxidation using water vapor progresses at a significantly faster rate than one using dry oxygen. Although dry oxidation provides more precise control of oxide thickness, the interface state density is lower for the wet oxide because hydrogen diffuses to the silicon–oxide interface and ties up the dangling bonds. The lowest fixed oxide charge and surface state density are obtained for the ⟨100⟩ orientation.

Initially, when the oxide layer is thin, the oxide growth rate is *reaction-rate-limited.* The oxide thickness d_{ox} follows the *linear law* [Appendix 8.1, Eq. (A8.1.22)]:

$$d_{\mathrm{ox}} = \alpha t \tag{8.1}$$

where α is the linear rate constant and t is the oxide growth time. The unit of α is μm/hr because from Eq. (8.1), $\alpha = d_{\mathrm{ox}}/t$ has the dimensions of length/time $= [L]/[T]$. For wet oxide at 1200°C, $\alpha = 0.72$ μm/hr. But when the oxide layer is thick, the growth rate is determined by the diffusion of oxidizing species through this layer, making it *transport-limited.* Then the oxide thickness obeys the parabolic law [Appendix 8.1, Eq. (A8.1.23)]:

$$d_{\mathrm{ox}} = \sqrt{\beta t} \tag{8.2}$$

where β is the parabolic rate constant. It is measured in $\mu m^2/hr$ because from Eq. (8.2), $\beta = d_{ox}^2/t$ has the dimensions of length^2/time $= [L^2]/[T]$. For wet oxide at 1200°C, $\beta = 14.4\ \mu m^2/hr$, which is $14.4/0.72 = 20$ times that for dry oxide at the same temperature, explaining its much faster growth rate.

During thermal oxidation of silicon, a layer of silicon itself is used up to form silicon dioxide. Consequently, an interface is formed between silicon and silicon dioxide. With the progress of oxidation, this interface moves into silicon. If the silicon is doped, P-type or N-type, the impurities present in the consumed silicon layer will be situated either in the oxide or in the silicon below. Redistribution of doping impurity takes place at the interface. This impurity redistribution is influenced by the diffusivity of impurity in the oxide and the rate of advancement of the interface with respect to the diffusion rate. The general rule is that acceptor impurities move into the oxide while donor impurities move into silicon; for example, boron accumulates in the newly grown SiO_2, whereas phosphorus, arsenic, and antimony tend to segregate in Si near the advancing Si–SiO_2 interface. This occurs because of the smaller solubility of boron in Si than in SiO_2, with the reverse being true for P, As, and Sb. This piling up of impurities decreases the P-type doping at and near the silicon surface. Also, N-type doping concentration increases near the surface region of silicon.

At equilibrium, the ratio of impurity concentration in silicon (C_{Si}) to the impurity concentration in silicon dioxide (C_{SiO_2}) is constant at the interface. This constant ratio is referred to as the *equilibrium segregation coefficient* (m) and is given by $m = C_{Si}/C_{SiO_2}$. The experimentally determined segregation coefficient is known as the *effective segregation coefficient.*

During oxidation of boron-doped silicon, the boron is depleted from the surface region of silicon. So, C_{Si} becomes smaller. This boron moves into the growing oxide layer making C_{SiO_2} larger. Hence, the ratio C_{Si}/C_{SiO_2} or m becomes less than unity. Reported m values for boron in the temperature range 850–1200°C lie between 0.1 and 1. Also, values for ⟨100⟩ orientation are greater than those for ⟨111⟩ orientation, owing to the smaller packing density of the ⟨100⟩ crystal. Hence, for the same C_{Si} value, C_{SiO_2} is larger for ⟨111⟩ orientation compared to the ⟨100⟩ case, and accordingly m is smaller for ⟨111⟩ orientation than for ⟨100⟩. Contrary to the behavior of boron, N-type doping impurities like phosphorus, arsenic and antimony tend to pile up in silicon. Therefore, C_{Si} increases and C_{SiO_2} decreases, making $m > 1$. Segregation coefficients for these dopants are approximately 10.

Exposure of the oxidation tube to HCl + dry O_2 before oxidation reduces the mobile ion contamination. During dry oxidation, about 1% HCl by volume in the gas stream is effective.

Example 8.1 Given that the linear and parabolic rate constants for dry oxidation of silicon at 1100°C are 0.3 $\mu m/hr$ and 0.027 $\mu m^2/hr$, respectively, find the silicon dioxide thickness for growth times: (a) 5 min and (b) 50 min. What will be the oxide thickness for wet oxidation at the same temperature if the corresponding rate constants are, respectively, 4.64 μm/hr and 0.51 μm^2/hr.

For dry oxidation of silicon, (a) growth time = 5 min = 5/60 = 0.0833 hr. For this short time span, the linear law is valid and the resulting oxide thickness = $d_{ox} = \alpha t =$ 0.3 × 0.0833 = 0.02499 μm = 0.02499 × 10^4 Å = 249.9 Å. (b) Growth time = 50 min = 50/60 = 0.833 hr. For this long time interval, the parabolic law is applicable and the oxide thickness = $d_{ox} = \sqrt{(\beta t)} = \sqrt{0.027 \times 0.833} = 0.14996$ μm = 0.14996 × 10^4 Å = 1499.6 Å.

For wet oxidation of silicon, (a) $d_{ox} = \alpha t = 4.64 \times 0.0833 = 0.3865$ μm = 0.3865 × 10^4 Å = 3865 Å. (b) $d_{ox} = \sqrt{\beta t} = \sqrt{0.51 \times 0.833} = 0.65179$ μm = 0.65179 × 10^4 Å = 6517.9 Å.

8.2.3. Thermal Diffusion Cycles

Thermal diffusion in silicon is necessary throughout the IGBT fabrication process. For accurate control and other advantages, if the impurity predeposition is performed by ion implantation (Section 8.2.4), its redistribution to achieve the required impurity profile is carried out by the drive-in process. During this stage, an oxide layer is also grown for the subsequent step so that the impurity drive-in (by thermal diffusion) and oxide growth (by thermal oxidation) are done together. Looking at Fig. 8.2, we find that there are several such drive-in and oxidation stages in the IGBT process sequence. Another example is phosphorus doping of polysilicon to lower its resistivity, which is done by phosphorus diffusion (or phosphorus predeposition by ion implantation followed by drive-in). The diffusion process in an IGBT (or any semiconductor device fabrication) is divided into two steps [16–18]:

(i) *Predeposition Cycle* This step involves the introduction of a carefully controlled amount of desired impurity into Si, whose profile is given by [Appendix 8.2, Eq. (A8.2.25)]

$$N(x) = N_0 \operatorname{erfc}\left(\frac{x}{2\sqrt{D_1 t_1}}\right) \tag{8.3}$$

where $N(x)$ is the concentration of the dopant at a distance x into the wafer, N_0 is the solid solubility of the dopant in Si at the predeposition temperature, D_1 is the diffusion coefficient of the dopant at the predeposition temperature, and t_1 is the time for which the predeposition is carried out.

(ii) *Drive-in Cycle* This step involves the redistribution of impurity to obtain the required junction depth and profile, expressed as [Appendix 8.2, Eq. (A8.2.36)]

$$N(x) = \frac{Q}{\sqrt{\pi D_2 t_2}} \exp\left(-\frac{x^2}{4D_2 t_2}\right) \tag{8.4}$$

where Q is the amount of dopant (in atoms/cm^3) introduced into Si during predeposition, given by [Appendix 8.2, Eq.(A8.2.39)]

$$Q = N_0\left(2\sqrt{\frac{D_1t_1}{\pi}}\right) \tag{8.5}$$

D_2 is the diffusion coefficient of the dopant at the drive-in temperature and t_2 is the drive-in time.

Example 8.2 During IGBT fabrication, boron predeposition is carried out at 950°C for 30 min on an N-type silicon wafer (1×10^{14} cm^{-3}) in an excessive boron environment. If the saturation concentration at 950°C is 3.8×10^{20} atoms/cm^3, calculate the boron junction depth. Now drive-in is to be performed at 1100°C to achieve a boron junction depth of 3 μm. Determine the doping cycle—that is, the time required for this purpose. What is the drive-in time at 1200°C?

The diffusivity is expressed as

$$D = D_0 \exp\left(-\frac{E}{kT}\right) \tag{E8.2.1}$$

where D_0 is the preexponential factor, E is the activation energy, k is Boltzmann's constant and T is the absolute temperature. For boron $D_0 = 0.76$ cm^2/sec and $E = 3.46$ eV. At 950°C = 1223 K, we have $D = 0.76\exp\{-3.46/(8.62 \times 10^{-5} \times 1223)\} = 4.24 \times 10^{-15}$ cm^2/sec. The boron junction depth x_j is determined from

$$N(x_j) = N_0 \operatorname{erfc}\left(-\frac{x_j}{2\sqrt{D_1t_1}}\right) \quad \text{or} \quad 1 \times 10^{14}$$

$$= 3.8 \times 10^{20} \operatorname{erfc}\left(-\frac{x_j}{2\sqrt{4.24 \times 10^{-15} \times 30 \times 60}}\right) \quad \text{or} \quad 2.63 \times 10^{-7}$$

$$= \operatorname{erfc}\left(-\frac{x_j}{5.5 \times 10^{-6}}\right) \tag{E8.2.2}$$

Using complementary error function table, $3.64 = -x_j/(5.5 \times 10^{-6})$ so that $x_j = 2.002 \times 10^{-5} = 0.2$ μm.

The amount of boron introduced during predeposition is

$$Q = N_0\left(2\sqrt{\frac{D_1t_1}{\pi}}\right) = 3.8 \times 10^{20}\left(2\sqrt{\frac{4.24 \times 10^{-15} \times 30 \times 60}{3.14}}\right)$$

$$= 1.185 \times 10^{15} \text{ atoms/cm}^2 \tag{E8.2.3}$$

$$N(x) = \frac{Q}{\sqrt{\pi D_2 t_2}} \exp\left(-\frac{x^2}{4D_2 t_2}\right) \quad \text{or} \quad 1 \times 10^{14}$$

$$= \frac{1.185 \times 10^{15}}{\sqrt{3.14 \times D_2 t_2}} \exp\left\{-\frac{(3 \times 10^{-4})^2}{4D_2 t_2}\right\}$$

$$\text{or} \quad 0.1495 = \frac{1}{\sqrt{D_2 t_2}} \exp\left(-\frac{2.25 \times 10^{-8}}{D_2 t_2}\right) \tag{E8.2.4}$$

Putting $D_2 t_2 = y$, we get

$$0.1495 = \frac{1}{\sqrt{y}} \exp\left(-\frac{2.25 \times 10^{-8}}{y}\right) \tag{E8.2.5}$$

Squaring both sides, we have

$$0.02235 = \frac{1}{y} \exp\left(-\frac{4.5 \times 10^{-8}}{y}\right) \tag{E8.2.6}$$

which is solved by trial and error giving $y = 1.8835 \times 10^{-9}$ cm^2. At 1100°C, $D = 0.76 \exp\{-3.46/(8.62 \times 10^{-5} \times 1373)\} = 1.5288 \times 10^{-13}$ cm^2/sec, yielding $y = D_2 t_2 = 1.5288 \times 10^{-13} \times t_2 = 1.8835 \times 10^{-9}$ so that $t_2 = 1.8835 \times 10^{-9}/1.5288 \times 10^{-13} = 12320.12$ sec $= 12320.12/3600 = 3.422$ hr. At 1200°C, $D = 0.76 \exp\{-3.46/(8.62 \times 10^{-5} \times 1473)\} = 1.1125 \times 10^{-12}$ cm^2/sec, yielding $y = D_2 t_2 = 1.1125 \times 10^{-12} \times t_2 = 1.8835 \times 10^{-9}$ so that $t_2 = 1.8835 \times 10^{-9}/1.1125 \times 10^{-12} = 1693.033$ sec $= 1693.033/3600 = 0.47$ hr $= 28.22$ min.

8.2.4 Ion Implantation

This process [19, 20], accelerates ions of a desired dopant by an electric field to a high energy (25–200 keV) and scans them across a wafer surface to produce a uniform predeposition. The controlling variables are the *dose* or the number of ions striking the wafer surface per unit area (5×10^{11} to more than 10^{16} cm^{-2}), determining the surface concentration, and the *energy* imparted to the ions, deciding the implantation depth. Typical ion-beam currents are ~ 1 mA, corresponding to a flux of 6.25×10^{15} singly charged ions per sec. To reduce the implantation time, implanters capable of ~ 15 mA are used. The average depth of implanted ions is termed the projected range R_p (Fig. 8.4). Distribution of ions about R_p is approximated by a Gaussian function with a standard deviation σ_p, called the projected straggle

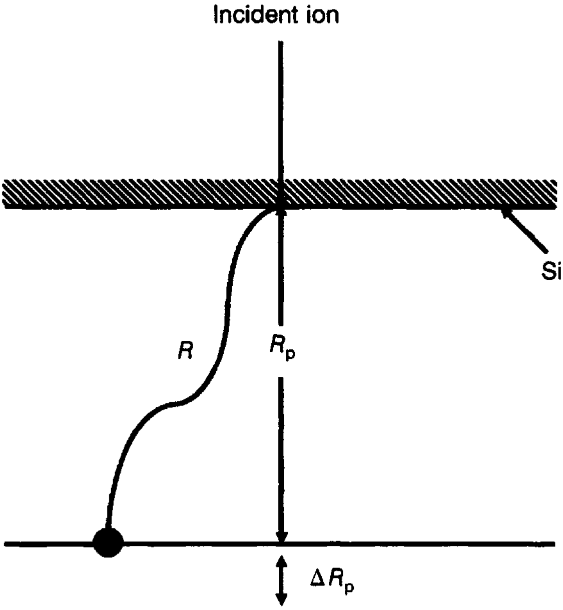


Figure 8.4 Schematic showing the ion range R, the projected range R_p, and the projected straggle ΔR_p.

ΔR_p. The impurity profile is expressed as

$$N(x) = N_0 \exp\left\{ -\frac{(x - R_p)^2}{2\sigma_p^2} \right\} \tag{8.6}$$

After ion implantation, the wafers are annealed at 900°C in nitrogen for a half hour to remove the lattice damage and place the dopant atoms at substitutional sites where they will be electrically active.

Example 8.3 For P base of IGBT, boron is implanted at 80 keV with 1×10^{14} atoms/cm^2, giving a projected range (R_p) of 0.2188 μm with projected standard deviation (σ_p) of 0.061 μm. This implantation is followed by annealing at 900°C for 1 hr. Calculate the boron concentrations at depths of 0.25 μm, 0.5 μm, and 0.75 μm from the surface. Also find the boron concentration at $x = R_p$.

The impurity profile after annealing is given by

$$N(x,t) = \frac{\phi}{\sqrt{2\pi\left(\sigma_p^2 + 2Dt\right)}} \exp\left\{ -\frac{(x - R_p)^2}{2\left(\sigma_p^2 + 2Dt\right)} \right\} \tag{E8.3.1}$$

where dose $\phi = 1 \times 10^{14}$ atoms/cm^2, projected range $R_p = 0.2188$ μm, projected standard deviation $\sigma_p = 0.061$ μm, and annealing time $t = 1$ hr $= 3600$ sec. At 900°C, diffusivity $D = 0.76\exp\{-3.46/(8.6174 \times 10^{-5} \times 1173)\} = 1.035 \times 10^{-15}$ cm^2/sec. Therefore, $\sigma_p^2 + 2Dt = (0.061 \times 10^{-4})^2 + 2 \times 1.035 \times 10^{-15} \times 3600 = 4.466 \times 10^{-11}$. For $x = 0.25$ μm, $x - R_p = (0.25 - 0.2188) \times 10^{-4} = 0.0312 \times 10^{-4}$. Hence,

$$N(0.6\ \mu\text{m}, 3600\ \text{sec}) = \frac{1 \times 10^{14}}{\sqrt{2 \times 3.14 \times 4.466 \times 10^{-11}}} \exp\left\{-\frac{9.7344 \times 10^{-22}}{2 \times 4.466 \times 10^{-11}}\right\}$$
$$= 5.97 \times 10^{18}\ \text{cm}^{-3} \tag{E8.3.2}$$

For $x = 0.5$ μm, $x - R_p = (0.5 - 0.2188) \times 10^{-4} = 0.2812 \times 10^{-4}$. Hence,

$$N(0.5\ \mu\text{m}, 3600\ \text{sec}) = \frac{1 \times 10^{14}}{\sqrt{2 \times 3.14 \times 4.466 \times 10^{-11}}} \exp\left\{-\frac{7.907 \times 10^{-10}}{2 \times 4.466 \times 10^{-11}}\right\}$$
$$= 8.54 \times 10^{14}\ \text{cm}^{-3} \tag{E8.3.3}$$

For $x = 0.75$ μm, $x - R_p = (0.75 - 0.2188) \times 10^{-4} = 0.5312 \times 10^{-4}$. Hence,

$$N(0.75\ \mu\text{m}, 3600\ \text{sec}) = \frac{1 \times 10^{14}}{\sqrt{2 \times 3.14 \times 4.466 \times 10^{-11}}} \exp\left\{-\frac{2.822 \times 10^{-9}}{2 \times 4.466 \times 10^{-11}}\right\}$$
$$= 1.13 \times 10^{5}\ \text{cm}^{-3} \tag{E8.3.4}$$

At $x = R_p$, $N(0.2188$ μm, 3600 sec$) = 5.97 \times 10^{18}$ cm^{-3}.

8.2.5 Photolithography [21-24]

Photolithography is the transfer of the image from the mask (a glass plate covered with an array of patterns) to the surface of the wafer through a sensitized material called the *photoresist*. It also includes the generation of the mask. Photoresist is a chemical formulation containing a light-sensitive material suspended in a solvent. It generally consists of a polymeric material that. upon sensitization by light, electron beam, or X rays, changes its chemical structure in such a way that the sensitized region either becomes more soluble (light-softened or positive photoresist) or becomes less soluble (light-hardened or negative photoresist) when immersed in a suitable solvent called the *developer*. In a positive photoresist, the molecular bonds are broken on illumination, whereas the molecules of a negative photoresist become cross-linked or polymerised on exposure. *Optical lithography* uses contact, proximity, or projection printing. The mask is fabricated by an electron-beam pattern generator, which, under computer control, exposes and places the pattern elements to form the chip image at 10 × scale. This image, called the *reticle*, serves as the object in a step-repeat camera, which steps the reduced image filling the required mask space. Another method of mask making is *laser beam scanning*. Multiple copies of the master mask produced are used for photolithographic work.

Optical lithography gives linewidths up to 0.4 μm. Improved resolution is provided by electron-beam, X-ray, and ion-beam lithographies. Due to the smaller wavelength (< 1 Å) of 10- to 50-keV electrons, much improved resolution is obtained by electron lithography. This technique has two subdivisions, namely, scanning and projection lithographies. The scanning method is further classified into raster and vector scanning.

Because of their pattern generating function, electron beam exposure machines are direct writing and mask fabrication for optical stepper, X-ray, and ion-beam lithographies. *Direct writing on wafer* is a mask-less process in which a high-resolution pattern is directly scanned on the resist-coated wafer by the moving electron beam. Advantages of direct writing include the facility to make different geometrical patterns on one wafer, the shorter time span from device design to testing and design modification, if necessary, and the high efficiency and manufacturing speed and the economic benefits. *X-ray lithography* gives resolution around 0.2 μm and placement accuracy ~ 0.3 μm. Moreover, contaminants like organic materials do not print as a defect. Depth of focus is also much larger when compared to optical printing. *Ion-beam lithography* has the advantage of higher resolution over electron lithography, due to the smaller degree of scattering on resist exposure, and the greater sensitivity of the resists to ions than to electrons. Ion lithography systems are either of the scanning focused-beam type or mask-beam type.

8.2.6 Chemical Vapor Deposition of Polycrystalline Silicon, Silicon Oxide, and Silicon Nitride

Chemical vapor deposition (CVD) is the formation of a nonvolatile solid film on a substrate by the reaction amongst vapor-phase chemicals containing the required ingredients, [25–27]. CVD is classified as atmospheric pressure (APCVD), low-pressure CVD (LPCVD), and plasma-enhanced CVD (PECVD).

Polysilicon (silicon with a short-range crystal structure) has a special importance in IGBT processing, where it is used as the gate electrode. It is formed if the deposition rate on a substrate is high, the substrate has no crystal structure, or the deposition temperature is lower than that for single-crystal growth.

$$SiH_4 + \text{Heat} \longrightarrow Si + 2H_2(\text{carrier gas} = N_2;\ \text{temperature} = 575\text{–}650°C)$$

20–30% silane, diluted in N_2, is fed to an LPCVD reactor at 0.2–1.0 torr (torr is a unit of pressure used in vacuum technology defined as 1 mm of mercury; 1 torr = 133.322 pascals). The properties of the polySi film are determined by the deposition temperature and pressure, and silane content.

A 0.5-μm-deep, heavily doped polysilicon film has a sheet resistivity of 20 ohms/square (Ω/□). (Ohms/square is the unit of sheet resistance or sheet resistivity. Sheet resistance of a substance is defined as the resistance of a square specimen of the material having unit thickness.)

The reaction for silicon dioxide deposition is

$$SiH_2 + 2O_2 \longrightarrow SiO_2 + 2H_2O(\text{carrier gas} = N_2;\ \text{temperature} = 200\text{–}500°C)$$

The process is conducted either at atmospheric pressure (APCVD) or reduced pressure (LPCVD).

Silicon nitride is deposited by either of the two pyrolytic processes:

(i) Atmospheric pressure CVD (APCVD) in which silane is reacted with ammonia at 700–900°C:

$$3SiH_4 + 4NH_3 \longrightarrow Si_3N_4 + 12H_2$$

(ii) Low-pressure CVD (LPCVD) in which dichlorosilane is reacted with ammonia under reduced pressure at 700–800°C:

$$3SiCl_2H_2 + 4NH_3 \longrightarrow Si_3N_4 + 6HCl + 6H_2$$

Low-pressure technique provides uniformity of film thickness and composition. It also gives high throughput of silicon wafers, which favors it for mass production. LPCVD is therefore superior to APCVD. The present trend in silicon nitride film deposition is toward hot-wall low-pressure reactors allowing close stacking of wafers but giving uniform film coverage.

8.2.7 Reactive Plasma Etching

Reactive plasma is a discharge in which the ionization and fragmentation of gases occurs, producing chemically active species, which are reactive both in the gaseous phase and with solid surfaces exposed to them. *Reactive plasma etching*, [28, 29], is the removal of material from a Si surface, unmasked by lithographic patterns to form volatile products. Halogen-based gases, such as CF_4, CCl_3F, and Cl_2, are generally used. In the plasma, CF_4 dissociates into fluorine atoms and fluorinated fragments. Neutral species, like atomic fluorine or atomic chlorine, produce etching while the accelerated ions bombarding the substrate knock away the exposed atoms, causing sputtering of the material (sputter-etching) and also making the substrate surface more active with respect to the neutral species. Reactions of atomic fluorine with Si and SiO_2 are

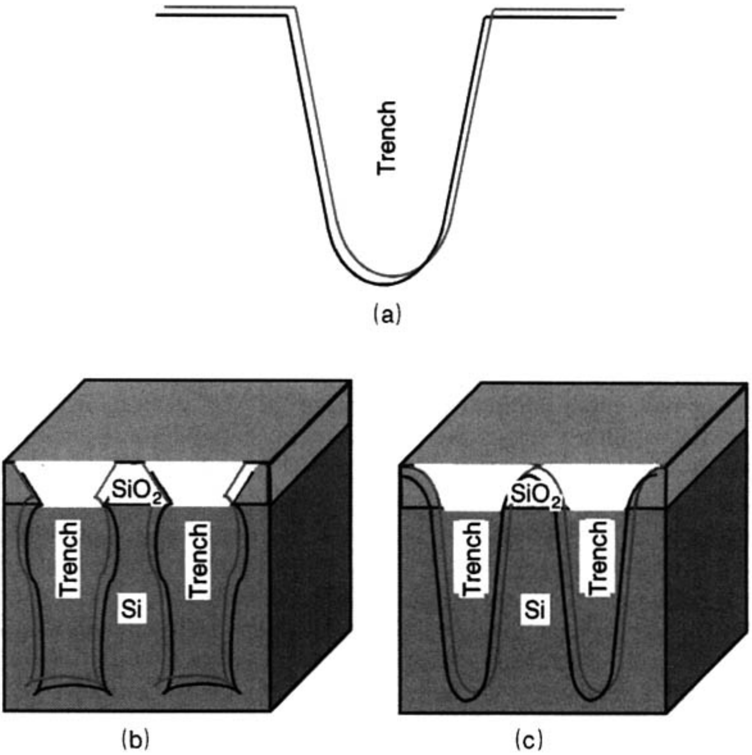
$$Si + 4F \longrightarrow SiF_4;\ SiO_2 + 4F \longrightarrow SiF_4 + O_2$$

Additional gases such as H_2 or O_2 provide the desired selectivity and edge profile. Selectivity implies faster etching of one film compared to another, under similar etching conditions. The etch rate of Si increases drastically by adding oxygen to CF_4 with a maximum etch rate at 12% O_2. Similarly, for SiO_2, the maximum etch rate is achieved at 20% O_2. This is ascribed to the falling recombination rate of F and CF_x. SiO_2 etching is not affected by the H_2 concentration, but Si etching slows down and eventually ceases at 40% H_2 concentration. The mechanisms responsible for this phenomenon are: the consumption of F atoms through their reaction with H_2 to form HF, together with the reaction of CF_4 with Si forming a carbon, or hydrocarbon polymer, on the Si surface, inhibiting further Si etching.

Wet etching is usually isotropic as it proceeds vertically downwards from the silicon surface and laterally underneath the masking layer. But dry etching can be made anisotropic by proper choice of the reactor conditions such as low pressures and directional electric fields, for more vigorous ionic bombardment. Hence, nearly vertical walls are etched in the material. But dry etching gives lower selectivity and is accompanied by substantial etching of the masking material. This requires accurate end-point detection, by monitoring the emission of characteristic light from the reaction, and terminating the etch process as soon as it is completed.

For gate etching in IGBTs, chlorine-based and bromine-based chemistries are used. Due to its higher etch selectivity of polySi compared to gate oxide, bromine-based chemistry gives superior performance when compared to chlorine-based etching. Prior to doing polySi etching, a fluorine-based plasma removes the thin native oxide. This avoids any micromasking formation during polySi etching. Further improvement in the selectivity of polySi, compared to gate oxide, is obtained by the removal of carbon-containing gases and materials from the reactor and the etching process.

Besides the etching of silicon oxide, nitride, polySi, and so on, an important application of plasma etching is that of trench formation, [30, 31]. An *ideal trench* has sloped walls with a rounded bottom (Fig. 8.5a). The sloped walls, produced by redeposition during etching, are necessary to avoid any void during the subsequent conformal deposition to refill the trench. The rounded bottom eliminates the electric field crowding due to sharp corners. Deep trench etching, required in trench-gate IGBTs, poses stringent requirements on etch rate, anisotropy, and selectivity. The etchant should have a high etch rate, be extremely selective to the masking material such as SiO_2, and yield anisotropic etching. Although fluorine chemistry possesses the desirable etch rate, the remaining criteria are not met. Fluorocarbon chemistry results in undercutting of mask material and is not suitable. Amongst the chlorine-based and bromine-based chemistries [32, 33], both of which have high etch rate and selectivity, slightly isotropic profiles are obtained with the chlorine-based chemistry. Use of a carbon-containing gas provides sidewall passivation. This passivation protects the sidewall from the lateral etching effects and local ion-enhanced trenching. The shape of the trench as



(a)

(b)

(c)

Figure 8.5 Vertical cross-sectional etching profile of trenches. (a) Ideal trench profile. (b) Trench made by conventional RIE. (c) Trench made by RIE with sidewall protection.

well as the smoothness of the sidewalls is determined by striking a balance between the etching and its prevention by the passivation layer. The temperature of the wafer strongly affects the sidewall shape. At a higher temperature, a smaller amount of passivation material is deposited, which favors more lateral etching. Fig. 8.5b shows the vertical cross-section of a trench without sidewall passivation to prevent lateral etching. Using sidewall etching, a better trench profile, shown in Fig. 8.5c, is realized.

Crucial factors in the etching technology include: high accuracy and selectivity, precision in controlling the trench shape, minimizing the damage to the electrical properties of the device, achieving high etching rate and consistency of the process.

8.2.8 Metallization

Metallization provides interconnections between contacts on the device and between the device and the outside world. Gate and interconnection metallization controls the switching speed of the device through the resistance of the interconnection runners. Their *RC* time constant, τ, is given by

$$\tau = \left(\frac{\rho}{d_{\text{polySi}}}\right)\left(\frac{L^2\varepsilon_{\text{ox}}}{d_{\text{ox}}}\right) \tag{8.7}$$

where ρ is the resistivity of polysilicon, d_{polySi} is its thickness, L is the length of the runner, ε_{ox} is the relative permittivity of silicon dioxide, and d_{ox} is the oxide thickness. Beside the device switching speed, metallization also affects the flat-band voltage (see Example 3.1) and hence threshold voltage (Section 3.2). Because of the ease of processing and patterning, low resistivity (2.7–3 $\mu\Omega$-cm), its ability to reduce native oxide on the silicon surface, surface smoothness, adherence, and mechanical stability, aluminum is the popular metal for the upper surface metallization of the device. Its drawbacks include: the resultant spikes due to the metallization process, which cause shorting of shallow junctions; poor step coverage; high contact resistance; electromigration or atomic transport at high current densities; and low melting point. It is deposited by thermal or electron-beam evaporation and by sputtering techniques. For Al, electromigration becomes severe at a current density of 10^5 A/cm^2. The addition of 2–3% Cu impedes electromigration, thereby augmenting the long-term high-current handling capability of Al without any significant effect on its resistivity. Cu is the most favored substitute material for Al. It is deposited by CVD. Obstacles to Cu include its vulnerability to corrosion and diffusion through SiO_2 warranting the use of a barrier layer like Si_3N_4.

For the bottom contact a tri-layer metallization scheme like Ti–Ni–Au, Ti–Ni–Ag, Cr–Ni–Au, and so on, is used. As already indicated in Section 8.1.1, the first metal layer (Ti or Cr) is the barrier metal, while the second layer (Ni) is for solderability. The third layer (Au or Ag) provides good wetting of the solder. Poor wetting produces voids in the solder leading to high electrical and thermal resistances.

Example 8.4 Find the *RC* time constant for a 0.5-μm-thick and 0.4-cm long runner of doped polysilicon (resistivity = 1×10^3 $\mu\Omega$-cm) on 0.7-μm-thick SiO_2.

The *RC* time constant is given by

$$\tau = \left(\frac{\rho}{d_{\text{polySi}}}\right)\left(\frac{L^2\varepsilon_{\text{ox}}}{d_{\text{ox}}}\right) + \left(\frac{1 \times 10^3 \times 10^{-6}}{0.5 \times 10^{-4}}\right)\left\{\frac{(0.4)^2 \times 8.854 \times 10^{-14} \times 3.9}{0.7 \times 10^{-4}}\right\}$$

$$= 1.5785 \times 10^{-8}\ \text{s} = 1.579 \times 10^{-8} \times 10^9\ \text{nsec} = 15.79\ \text{nsec} \tag{E8.4.1}$$

8.2.9 Electron Irradiation

This is performed at typical energies of 3 MeV and doses up to 16 Mrad after the metallization—for example, when chips are mounted on packages. The electron beam is made to impinge directly on the emitter surface. The electron beam dose is controlled by changing the beam current. The irradiation is carried out at room temperature. The oxide charge is annealed out by baking at 140°C until the threshold voltage has recovered to a predecided value (e.g., 3 V). Any overdose is recoverable by heating the devices at 400°C. Electrical parameters like turn-off time, forward drop, and leakage current are measured. For illustrative purposes, the turn-off time decreased from 20 μsec to 200 nsec with the forward drop rising from 1.8 V to 5 V at 10-A collector current, but the leakage current remained below 1 μA. Overall, it affords a simple and clean process.

The relationship between the preirradiation lifetime τ_0 and postirradiation lifetime τ after a dose ϕ is

$$\frac{1}{\tau} = \frac{1}{\tau_0} + K\phi \tag{8.8}$$

where the radiation damage coefficient K depends upon type and energy of radiation, as well as upon the wafer resistivity and temperature. Therefore, the radiation dose must be tailored for the individual device design in hand.

8.2.10 Proton Irradiation

The useful property of proton radiation [34] is that its penetration depth in silicon can be confined within the silicon wafer thickness. For example, the range of 3-MeV protons in Si is 100 μm, while that of equal energy electrons is 6000 μm. Since maximum radiation damage occurs toward the end of the range, this property enables us to create defects locally in a narrow, well-defined band near the junction, at a known silicon depth below the surface, where lifetime reduction is desired. In the remaining wafer, the damage is relatively less. Thus switching speed of the IGBT is enhanced without sacrificing other parameters like forward voltage drop and leakage current. The main disadvantage of proton irradiation is the necessity of performing the process under vacuum conditions, thereby making this technique costly.

8.2.11 He Implantation

A vertical control of the damaged region down to 10 μm is possible with proton irradiation. An efficient method of localized lifetime control in IGBTs is based on the formation of void layers by low energy (50–80 keV) and high dose (1×10^{16} to 6×10^{16} /cm^2) He, or α implantation in the N-buffer layer of IGBT followed by annealing at 900°C for 1 hr [35]. These voids are well-controlled in depth (layers < 100 nm thick), with sharp lateral definition. They introduce two well-defined trap levels in the middle of the energy gap, one at 0.55 eV from the conduction band for electrons and the other at 0.53 eV from the valence band for holes. Comparison of IGBTs with voids in the buffer layer and IGBTs with unlocalized recombination centers has revealed that IGBTs with voids have a smaller forward conduction drop than those with unlocalized lifetime control, for the same turn-off time values. This improvement is enhanced for fast-switching IGBTs having turn-off times below 200 nsec, resulting in superior IGBT performance.

Lifetime engineering by formation of voids in Si through He implantation is a powerful technique to increase carrier recombination in IGBTs for controlling its switching performance. This method is very similar to the standard microelectronics fabrication procedure, not requiring any special equipment. Mapping measurements on 150-mm wafers have shown good uniformity and suitability for an industrial environment.

8.2.12 Packaging

Like the other discrete devices and IC chips, IGBT packaging is a broad subject encompassing the steps from the preassembly wafer preparation to the package fabrication techniques [36]. The package serves the threefold purpose of providing the electrical connection to the IGBT chip, protecting it from mechanical and environmental stresses, and providing the thermal path for heat dissipation. The package frequently dominates the overall cost, performance, and reliability of the device.

Chip Separation. Generally called "wafer scribing," the three commonly used methods are: diamond scribing using a tool with precisely shaped diamond embedded in it, laser scribing by producing a series of holes along the line, and sawing with the help of rotating blades.

Die Bonding (or Die-Attach). Eutectic die bonding metallurgically fixes the die to the metal leadframe or ceramic substrate, usually coated with Ag or Au. The backside of the die is metallized by a layer of gold to make it wettable by the die bonding solder preform (Au-98%, Si-2%) which dissolves the silicon at the eutectic temperature (370°C). When placing the preform and then the die on the leadframe/substrate, and then heating them together, to the eutectic temperature, a bond is formed. On cooling, the bond is completed when the composite freezes. The eutectic process is contamination-free, providing excellent shear strength, but it exposes the die to thermal stresses and is difficult to automate.

Wire Bonding. Gold or aluminum wires are used. Gold is preferred for its corrosion resistance and higher current-carrying capability. Gold wires are attached by thermo-compression bonding (using both heat and pressure) while aluminum wires are attached by ultrasonic bonding using a pulse of ultrasonic energy.

Although these wires are necessary for device operation, they introduce parasitic resistance and capacitance, degrading the performance.

Packages. Two types of packages are commonly used, namely, ceramic or plastic, with metal leads. In a hermetic-ceramic package, the IGBT chip resides in a vacuum-tight enclosure, completely uninfluenced by the external environment. Ceramic packages are more expensive, provide more reliable operation with hermetic sealing, and offer higher thermal conductivity (especially with beryllia or aluminum nitride) and thermal expansion coefficient, matching that of silicon. These packages are preferred for high-performance applications where the cost disadvantage can be tolerated. In a plastic package, the chip is encapsulated with resin materials. Effects of the outside ambient become significant over extended periods of time by penetration through the plastic. Due to the automated batch handling, plastic packages are cheaper, but they are less reliable, and are employed where hermeticity is not required. However, in postmolded plastic packaging, the weak wire bonds are removed by the injection molding process; and shrinkage after molding keeps these bonds in compressive loading, making them more reliable. The polymers used are novolac epoxies (produced by the reaction of novolac resin with epichlorohydrin and a base) and silicones, comprising a polysiloxane backbone with different organic substituents at the Si atom.

8.3 PROCESS INTEGRATION AND SIMULATION

Analytical determination of the impurity dopant profiles is extremely difficult. Moreover, analytical approximations for doping profiles typically do not adequately reflect results of fabrication processing, especially for devices with submicron dimensions. But theoretical prediction of processing variables is essential because their experimental determination is time-consuming and expensive. Sophisticated computer software packages have been developed to model and optimize the device structure without recourse to extensive experimentation. Computer-aided design (CAD) tools constitute an integral constituent of the process line. These programs expedite the development of innovative structures. In these programs, the mathematical techniques are a combination of numerical methods and analytical expressions. Accurate modeling (e.g., using the more complicated diffusion equation), instead of the simple Fick's law, invariably involves equations solvable only by the numerical approach. Therefore, these programs are a compromise between the accuracy, complexity, and computation times. Presently, a number of commercial two-dimensional process simulators are available, mostly for UNIX-based workstations. Normally they require tens of Mbytes of memory even for modest size meshes. Increasing performance and wide spread availability of PCs and compatibles encourage the development of software tools that can be used for two-dimensional modeling of semiconductor processes with a low memory capacity and high speed of computation. Nevertheless, these are powerful modeling tools for industrial semiconductor process/design. The algorithm is based on the finite-difference formulation on rectangular, auto-adjusting meshes.

Table 8.1 Process Characterizing Parameters

Serial No.	Unit Process Step	Process Parameter
1	Epitaxy	Type of impurity (B, P, As, Sb); impurity concentration in the ambient; time and temperature of growth; growth rate; silane (SiH_4) partial pressure.
2	Thermal oxidation	Process temperature, time, ambient pressure, nature of ambient (dry, wet, chlorine), percentage and source of chlorine (for chlorine ambient such as H_2O–HCl or O_2–C_2HCl_3).
3	Thermal diffusion	**Predeposition:** Temperature, time, ambient gas pressure, type of impurity (B, P, As, Sb); impurity concentration in the ambient; and ambient (e.g., main gas flow, N_2, carrier gas, N_2, and reacting gas, O_2 for $POCl_3$). **Drive-in (annealing):** Temperature, time, ambient pressure, and ambient (O_2 for dry oxidation, O_2 bubbled through hot water for wet oxidation, or only N_2 for drive-in without oxidation).
4	Ion implantation	Energy, dose, Type of impurity (B, P, As, Sb), angle of susceptor.
5	Deposition	Material (oxide/nitride/polysilicon/aluminum), temperature, pressure, grain size, impurity (B, P, As, Sb), impurity concentration, thickness of the layer, deposition rate and time.
6	Etching	Material to be etched (oxide/nitride/polysilicon/aluminum), thickness, etching time.

The program input file describes the processing schedule. This file contains the sequence of unit process steps and the respective process parameter, as defined in Table. 8.1. The chart of temperatures, times, ambients, and other parameters for epitaxial growth, oxidation, diffusion, ion implantation, deposition, and etching is carefully laid out. The program output is a one-dimensional profile in the bulk silicon or some overlying layers like polysilicon. Two-dimensional cross-sectional views of the device, as the process progresses, are also obtainable.

Although the protocols for writing the fabrication steps differ widely amongst the various software packages, a general description of the process simulation procedures will be in order. The most convenient way of starting a new project is to copy a proper existing project. A new project will be created.

The input file of the process simulation program contains several lines explaining the problem to the computer. First, the analysis is given a title name for identification such as

Title: IGBT Process Simulation

Then the starting silicon wafer specifications are entered, including the doping element, the concentration, and the silicon crystal orientation, for example,

Substrate Definition: Element = phosphorus,

Concentration = 1×10^{14} cm^{-3}, Orientation = $\langle 100 \rangle$

Definition of the grid structure follows this step. Domain and mesh parameters are decided. The grid is divided into high-resolution and low-resolution regions. The former is situated near the surface of the silicon wafer where the device is to be fabricated. This is the region in which the different junctions lie. The latter begins and extends farther below the surface. Not much activity takes place in this region so that the grid spacing need not be small. So, it is necessary to lay down the grid spacing in the different regions, as well as their locations. Number of mesh nodes in x direction (along the surface) and y direction (into the depth of the domain) is defined. A greater number of mesh nodes gives a higher computational accuracy at the expense of a larger CPU time. Domain size in x direction (in microns) should cover a region near edges of all the masks where the two-dimensionality takes place. Domain size in y direction (microns) should be deep enough to cover the maximum expected depth of the implanted or deposited dopant penetration. Domain size in z direction is decided by device width. A uniform mesh or a mesh exponentially condensed in the origin of coordinates is laid out. An example of grid definition is:

Grid Definition: Grid spacing in high-resolution region = 0.001 μm,

Depth of this region = 2 μm, Total simulation depth = 5 μm

Then the program assigns the grid spacing as 0.001 μm in the high-resolution region, and it automatically calculates the grid spacing in the remaining region.

The fabrication process includes processing steps such as ion implantation or surface deposition (arsenic, boron, or phosphorus) with subsequent annealing under oxidizing or inert ambient. Resulting doping profiles may be used in a straightforward manner to generate the entire structure of a semiconductor. The input file contains directives and parameters. All the directives are of two types: basic directives and model parameter directives. Basic directives include computational domain and mesh parameters, substrate parameters, numerical solution control, boron, phosphorus or arsenic deposition, boron, phosphorus or arsenic implantation, oxidation, annealing, and epilayer formation parameters. Model parameter directives are: diffusivity of boron, arsenic, and phosphorus, oxidation-enhanced diffusion, dry and wet oxidation kinetic constants, local oxidation "bird's beak" formula parameters, and segregation parameters (coefficient, activation energy, etc.).

The physical model adopted describes the diffusion process for up to three interacting charged impurities in a two-dimensional domain with moving oxide boundary and impurity segregation at the Si–SiO_2 interface. The oxidation of silicon is accompanied by the segregation—in other words, a leap in the impurity concentration at the moving Si/SiO_2 interface. The segregation causes an impurity flux density at the interface. For phosphorus and arsenic, the segregation coefficient is large (about 100) and usually close to the equilibrium value so that the impurity may be considered to be completely pushed into the silicon. The total impurity dose within the semiconductor and oxide is conserved. For the solution of several coupled diffusion equations the finite-difference equations of each impurity are solved sequentially, with initial values of impurity concentrations taken from the previous iteration or previous time step. Iterations continue until the solution for all impurities converges to a given accuracy.

To give an example, the oxidation step is specified as follows:

Step = Oxidation, Temperature = 1100°C,

Time = 100 min, Ambient = Dry O_2

Type of annealing atmosphere (dry or wet oxygen), pressure of the oxidizing ambient (oxygen or vapor), 1 atmosphere or higher, position of the oxidation mask, and so on, are given. Similarly, implantation and annealing steps are written as:

Step = Implantation, Element = Boron, Dose = 1×10^{15} cm^{-2}, Energy = 60 keV

Step = Annealing, Temperature = 900°C, Time = 60 min, Ambient = N_2

Drive-in and oxidation step is given as:

Step = Oxidation, Temperature = 1200°C, Time = 30 min, Ambient = Dry O_2

Step = Oxidation, Temperature = 1200°C, Time = 60 min, Ambient = Wet O_2

Step = Oxidation, Temperature = 1200°C, Time = 30 min, Ambient = Dry O_2

Polysilicon deposition is:

Step = Deposition, Material = Polysilicon, Temperature = 600 °C

Thickness = 5000 Å, Pressure = 0.8 atmosphere

Table 8.2 Effects and Models in Technological Simulators

Serial No.	Process	Effects and Models
1	Epitaxy	Epitaxial growth model, doping model (including transient effects for establishment of steady-state deposition process and time-variation of gas phase composition; trapping of the dopant atoms in incorporation sites, and autodoping effects).
2	Thermal oxidation	Semiempirical model, traditional Deal–Grove model [37, 38], influence of strain on molecular diffusion, empirical enhancement factor model for increase of oxidation rate during chlorine oxidation, heavy doping effects.
3	Diffusion	Empirical model, nonlinear diffusion, vacancy-point defect interaction-based model, dependence of diffusion coefficient on oxidation rate, concentration and oxidation-enhanced diffusion, emitter push-out and phosphorus pile-up effects, dopant segregation across material interfaces.
4	Ion implantation	Description of scattering of ions by changes in statistical energy distribution function, Monte-Carlo simulation, nuclear and electronic scattering effects.
5	Deposition	Simulation using Arrhenius relation, grain growth model, dopant segregation at grain/grain boundary.

After defining all the steps of the process, the format in which the output results are desired is explained. The manner in which the results are to be plotted and printed is described.

The models (Table 8.2) included in semiconductor technological simulators are undergoing continued revision and updating in view of recent developments. Moreover, this understanding of several physical and chemical processes is far from complete. Also, the process times must be made reasonable, and therefore approximations are unavoidable. Compounding of errors due to approximations in the unit process steps may sometimes lead to grossly erroneous results. Therefore, computer simulations must be supplemented by experimental investigations and precise measurements for design validation.

Popular commercial process simulation programs include (a) one-dimensional process simulators like SUPREM and (b) two-dimensional process simulators such as SUPRA and DIOS-ISE. These programs enable the

Table 8.3 A Typical Detailed Process and Assembly Sequence of IGBT Fabrication for Illustration

Serial No.	Process Step
1	Silicon wafers are taken (dopant = B, concentration = 1×10^{19} cm^{-3}, orientation = ⟨100⟩, thickness = 300 μm)
2	Epitaxy Layer I: Dopant = P, concentration = 1×10^{16} cm^{-3}, thickness = 10 μm Layer II: Dopant = P, concentration = 1×10^{14} cm^{-3}, thickness = 100 μm.
3	Field Oxidation Temperature = 1100°C, time = 30 min, dry O_2 Temperature = 1100°C, time = 100 min, wet O_2 Temperature = 1100°C, time = 30 min, dry O_2 Thickness = 0.8 μm
4	Patterning of the windows for deep P^+ implantation (Mask I# P^+-implantation), oxide etching and phoresist stripping
5	Growth of scattering oxide Temperature = 900°C, time = 120 min, dry O_2 Thickness = 300 Å
6	P^+ Implantation (boron, dose = 1×10^{16} cm^{-2}, energy = 60 keV)
7	Annealing Temperature = 900°C, time = 30 min, ambient = N_2.
8	Scattering oxide etching
9	Boron drive-in and oxidation Temperature = 1100°C, time = 30 min, dry O_2 Temperature = 1100°C, time = 60 min, wet O_2 Temperature = 1100°C, time = 30 min, dry O_2 Thickness = 0.65 μm
10	Patterning of the active area (Mask II # active), oxide etching and phoresist stripping.
11	Gate oxidation Temperature = 1000°C, dry O_2, thickness = 100 nm
12	Polysilicon deposition (temperature = 620°C, silane = 100%, time = 25 min, thickness = 560 nm)
13	Polysilicon doping (dopant = P, concentration, 1×10^{20} cm^{-3}, temperature = 950°C, time = 30 min)
14	Patterning of polysilicon for P-base diffusion (Mask III# Poly), oxide etching, and phoresist stripping
15	Growth of scattering oxide Temperature = 900°C, time = 120 min, dry O_2 Thickness = 300 Å

Table 8.3 (*Continued*)

Serial No.	Process Step
16	P^+ implantation (boron, dose = 1×10^{14} cm^{-2}, energy = 80 keV)
17	Annealing Temperature = 900°C, time = 30 min, ambient = N_2
18	Scattering oxide etching
19	Boron drive-in and oxidation Temperature = 1100°C, time = 30 min, dry O_2 Temperature = 1100°C, time = 60 min, wet O_2 Temperature = 1100°C, time = 30 min, dry O_2
20	Patterning of regions for emitter diffusion (Mask IV# N-Implantation), oxide etching, and phoresist stripping
21	Growth of scattering oxide Temperature = 900°C, time = 120 min, dry O_2 Thickness = 300 Å
22	N^+ implantation (phosphorus, dose = 1×10^{15} cm^{-2}, energy = 50 keV)
23	Annealing Temperature = 900°C, time = 30 min, ambient = N_2
24	Scattering oxide etching
25	Phosphorus drive-in and oxidation Temperature = 1000°C, time = 30 min, dry O_2
26	Growth of CVD oxide (thickness = 900 nm)
27	Reflow Temperature = 1050°C, time = 20 min, ambient = N_2
28	Patterning of contact regions (Mask V# Contact), oxide etching and photoresist stripping
29	Aluminum deposition (thickness = 6 μm)
30	Patterning of metal (Mask VI# Metal), metal etching and photoresist stripping
31	Al sintering Temperature = 450°C, time = 30 min, ambient = N_2
32	Back metallization (Ti–Ni–Au)
33	Testing
34	Wafer dicing for chip separation
35	Chip mounting on package base plate by solder preform (eutecting bonding)
36	Wire bonding
37	Electron irradiation and turn-off time adjustment
38	Final testing and encapsulation

device engineer to visualize the process from the starting wafer to the finished device. High-temperature diffusion, thermal oxidation, ion implantation, chemical vapor deposition, etching, and other unit processes are studied and the influence of a step performed later in the process sequence on an earlier step is analyzed. Tolerance in the process steps is also studied. Based on an extensive process simulation, the time, temperatures, and ambient for various steps are finalized and the detailed process sequence for IGBT fabrication is prepared (Table 8.3.

REVIEW EXERCISES

8.1 Explain how the fabrication of IGBT involves a union of MOS and bipolar technologies.

8.2 Name the photomasking steps in an IGBT fabrication sequence, and identify those you consider most critical, and why?

8.3 What metal layer is deposited on the upper surface of the IGBT? What desirable properties should the bottom metal layer possess?

8.4 Explain how is the trench–gate IGBT structure realized? What is the minimum trench width defined by the processing limits?

8.5 How is the switching speed of IGBT increased? What performance parameters are degraded by this process step?

8.6 The term "epitaxy" is the transliteration of two Greek words *epi* and *taxis*. What are their meanings in English? Write down the chemical reactions for epitaxial deposition of silicon. What competing reaction removes the silicon from the substrate? How are the dopants introduced during the epitaxial process?

8.7 When is the silicon dioxide growth rate reaction-rate limited and when it is transport-limited? Differentiate between dry and wet oxides in terms of growth rates and interface densities. Which silicon orientation gives the lowest surface state density? Name the dopants that segregate into the SiO_2 interface, and also those which pile up in Si.

8.8 What are the two steps by which thermal diffusion is performed for device fabrication? Which step gives the complementary error function profile and which one Gaussian profile? Write down the relevant equations.

8.9 Define the projected range and straggle of the ion implanted ions. What is the significance of post-implantation annealing?

8.10 Explain the terms "pholithography" and "photoresist." Name two methods of mask fabrication. What is meant by "direct writing on wafer." Give the advantages of electron, X-ray, and ion lithographies over the conventional optical lithography.

8.11 What do you understand by chemical vapor deposition? What are the three main types of CVD? Write down the chemical reactions and equations for deposition of polySi, SiO_2, and Si_3N_4, and mention the required conditions.

8.12 What roles are played by the neutral species (atomic fluorine or atomic chlorine) and the accelerated ions during plasma etching? What is meant by etch selectivity? How does the incorporation of oxygen or hydrogen in the etching mixture influence the etching of Si and SiO_2. Explain giving reasons.

8.13 List the desirable properties of an ideal trench, justifying their necessity? Mention the etchant characteristics required for excavating the deep trench required for trench-gate IGBT fabrication. Which chemistry is preferred for such trench formation?

8.14 How does the gate interconnection metallization affect the device speed? Argue for and against the use of aluminum as a versatile contact. Cite one alternative for Si, mentioning its advantages and limitations. Write down a tri-layer metallization scheme for the back contact of IGBT.

8.15 How are electron irradiation and proton irradiation performed for making fast IGBTs. Point out the advantages and disadvantages of these methods.

8.16 Name the two types of packages used for IGBTs, and describe their salient features. Explain "eutectic bonding."

8.17 List four process parameters that characterize: (a) thermal oxidation, (b) diffusion, and (c) ion implantation. Mention one uni-dimensional and one two-dimensional technology simulation package.

8.18 "Analytical formulations are adequate for a first-order analysis; but to account for second-order effects, it is necessary to take the help of numerical simulations." Do you agree with this statement? If so, how?

8.19 Determine the thickness of dry SiO_2 grown at 1100°C for (a) 7 min and (b) 62 min. The linear rate constant for dry oxidation of silicon at 1100°C is 0.3 μm/hr and parabolic rate constant is 0.027 μm^2/hr.

8.20 Estimate the silicon dioxide thickness for wet oxidation of silicon at 1100°C for (a) 2 min and (b) 109 min if the linear and parabolic rate constants for wet oxidation at 1100°C are, respectively, 4.64 μm/hr and 0.51 μm^2/hr.

8.21 During an IGBT fabrication process, boron predeposition is performed at 950°C for 30 min on an N-type silicon wafer (dopant density $= 2 \times 10^{14}$ cm^{-3}) using an infinite boron source. Then boron drive-in is performed at 1100°C to achieve a junction depth of 3 μm. What is the doping cycle—that is, the time required for achieving the required junction depth. The solid solubility of boron at 950°C is 3.8×10^{20} atoms/cm^3.

8.22 Calculate the *RC* time constant for a doped polysilicon line of thickness 0.25 μm and length 0.5 cm on 1-μm thick silicon dioxide, given that resistivity of doped polysilicon is 1×10^{-3} Ω-cm.

REFERENCES

1. C. Y. Chang and S. M. Sze, *VLSI Technology*, McGraw-Hill, New York, 1996.
2. S. M. Sze (Ed.), *VLSI Technology*, McGraw-Hill, New York, 1988.
3. S. K. Ghandhi, *VLSI Fabrication Principles: Silicon and Gallium Arsenide*, Wiley, New York, 1983.
4. S. P. Murarka and M. C. Peckerar, *Electronic Materials: Science and Technology*, Academic Press, Boston, 1989.
5. H. H. Lee, *Fundamentals of Microelectronics Processing*, McGraw-Hill, New York, 1990.
6. P. E. Gise and R. Blanchard, *Semiconductor and Integrated Circuit Fabrication Techniques*, Reston, Reston, 1979.
7. T. Ohmi, and T. Shibata, Developing a Fully Automated Closed Wafer Manufacturing System, *Micro-contamination*, Vol. 8, 1990, p. 25.
8 H. Iwamoto, H. Haruguchi, Y. Tomomatsu, J. F. Donlon, and E. R. Motto, A New Punchthrough IGBT Having a New N-Buffer Layer, *IEEE Trans. Ind. Appl.*, Vol. 38, January/February, 2002, pp. 168–174.
9. H. R. Chang, B. J. Baliga, J. W. Kretchmer, and P. A. Piacente, Insulated Gate Bipolar Transistor (IGBT) with a Trench Gate Structure, *International Electron Devices Meeting (IEDM)*, December 6–9, 1987, Washington, D.C., *Technical Digest*, IEEE, New York, pp. 674–677.
10. J. Bloem and L. J. Gilling, Epitaxial Growth by Chemical Vapour Deposition, in *VLSI Electronics*, N. G. Einspruch and H. Huff (Eds.), Vol. 12, Academic Press, Orlando, FL, 1985, p. 89.
11. B. J. Baliga, Defect Control During Silicon Epitaxial Growth Using Dichlorosilane, *J. Electrochem.Soc.*, Vol. 129, No. 5, May 1982, pp. 1078–1084.
12. J. R. Ligenza and W. G. Spitzer, The Mechanisms for Silicon Oxidation in Steam and Oxygen, *J. Phys. Chem. Solids*, Vol. 14 , 1960, p. 131.
13. W. A. Tiller, On the Kinetics of the Thermal Oxidation of Silicon, I, *J. Electrochem. Soc.*, Vol. 127, No.3, March 1980, pp. 619-624; II, Vol. 127, No. 3, March 1980, pp. 625–632; III, Vol. 128, No. 3, March 1981, pp. 689–697.
14. R. B. Fair, Oxidation, Impurity Diffusion, and Defect Growth in Silicon—An Overview, *J. Electrochem. Soc.*, Vol. 128, No.6, June 1981, pp. 1360–1368.
15. E. A. Irene, H. Z. Massoud, and E. Tierney, Silicon Oxidation Studies: Silicon Orientation Effects on Thermal Oxidation, *J. Electrochem. Soc.*, Vol. 133, No. 6, June 1986, pp. 1253–1256.
16. W. E. Beadle, J. C. C. Tsai, and R. D. Plummer, *Quick Reference Manual for Silicon Integrated Circuit Technology*, Wiley, New York, 1985.
17. U. Gosele, Current Understanding of Diffusion Mechanisms in Silicon, *Semiconductor Silicon*, 1986, H. R. Huff, T. Abe, and B. Kolbesen (Eds.), Electrochemical Society, Pennington, NJ, 1986, p. 541.
18. R. B. Fair, Concentration Profiles of Diffused Dopants in Silicon, in F. F. Y. Wang (Ed.), *Impurity Doping Processes in Silicon*, North-Holland, New York, 1981, Chapter 7.
19. W. Shockley, *Forming Semiconductor Devices by Ionic Bombardment*, US Patent 2787564.
20. M. I. Current, Current Status of Ion Implantation Equipment and Techniques for Semiconductor Fabrication, *Nucl. Inst. Methods*, Vol. B6, 1985, p. 9.
21. E. G. Cromer, Mask Aligners and Steppers for Precision Microlithography, *Solid State Technol.*, Vol. 36, 1993, p. 23.
22. K. Meissner, W. Haug, S. Silverman, and S. Sonchik, Electron-Beam Proximity Printing of Half-Micron Devices, *J. Vac. Sci. Technol.*, Vol. B7, No. 6, November/December 1989, pp. 1443–1447.

23. D. Fleming, J. R. Maldonado, and M. Neisser, Prospects for X-Ray Lithography, *J. Vac. Sci. Technol.*, Vol. B10, No. 6, November/December 1992, pp. 2511–2515.
24. J. L.Bartelt, Masked Ion Beam Lithography: An Emerging Technology, *Solid State Tech.*, Vol. 29, No. 5, May 1986, pp. 215–220.
25. T. Kamins, *Polycrystalline Silicon for Integrated Circuit Application*, Kluwer, Academic Publishers, Boston, 1988.
26. C. Cobianu and C. Pavelescu, Silane Oxidation Study: Analysis of Data for SiO_2 Films Deposited by Low Temperature Chemical Vapour Deposition, *Thin Solid Films*, Vol. 117, July 20, 1984, pp. 211–216.
27. T. Makino, Composition and Structure Control by Source Gas Ratio in LPCVD SiN_x, *J.Electrochem. Soc.*, Vol. 130, No. 2, February 1983, pp. 450–455.
28. R. A. Morgan, *Plasma Etching in Semiconductor Fabrication*, Elsevier, New York, 1985.
29. G. S. Oehrlein and J. F. Rembetski, Plasma-Based Dry Etching Techniques in the Silicon Integrated Circuit Technology, *IBM J. Res. Dev.*, Vol. 36, 1992, p. 140.
30. Y.-J. Lii and J. Jorné, Recombination During Deep Trench Etching, *J.Electrochem. Soc.*, Vol. 137, No. 9, September 1990, pp. 2837–2845.
31. H. Hübner, Calculations on Deposition and Redeposition in Plasma Etch Processes, *J. Electrochem. Soc.*, Vol. 139, No. 11, November 1992, pp. 3302–3309.
32. H. Crazzolara and N. Gellrich, Profile Control Possibilities for a Trench Etch Process Based on Chlorine Chemistry, *J. Electrochem. Soc.*, Vol. 137, No. 2, February 1990, pp. 708–717.
33. G. Wöhl, A. Weisheit, I. Flohr, and M. Bottcher, Trench Etching Using a $CBrF_3$ Plasma and Its Study by Optical Emission Spectroscopy, *Vacuum*, Vol. 42, No. 14, 1991, pp. 905–910.
34. M. Saggio, V. Raineri, R. Letor and F. Frisina, Innovative Localized Lifetime Control in High-Speed IGBTs, *IEEE Electron Device Lett.*, Vol. 18, No. 7, July 1997, pp. 333–335.
35. V.Raineri, M.Saggio, F. Frisina, and E.Rimini, Voids in Silicon Power Devices, *Solid State Electron.*, Vol. 42, No. 12, December 1998, pp. 2295–2301.
36. (a) T. Ohsaki, Electronic Packaging in the 1990s—A Perspective from Asia, *IEEE Trans. Components, Hybrids, Manuf. Technol.*, Vol. 14, No. 2, June 1991, pp 254–261. (b) R. R. Tummala, Electronic Packaging in the 1990s—A Perspective from America, *IEEE Trans. Components, Hybrids, Manuf. Technol.*, Vol. 14, No. 2, June 1991, pp. 262–271. (c) H. Wessely, O. Fritz, M. Horn, P. Klimke, W. Koschnick, and K.-H. Schmidt, Electronic Packaging in the 1990s—A Perspective from Europe, *IEEE Trans. Components, Hybrids, Manuf. Technol.*, Vol. 14, No. 2, February 1991, pp. 272–284.
37. B. E. Deal, The Oxidation of Silicon in Dry Oxygen, Wet Oxygen and Steam, *J. Electrochem. Soc.*, Vol. 110, No. 6, June 1963, pp. 527–533.
38. B. E. Deal and A. S. Grove, General Relationship for the Thermal Oxidation of Silicon, *J. Appl. Phys.*, Vol. 36, No. 12, December 1965, pp. 3770–3778.

APPENDIX 8.1: THERMAL OXIDATION OF SILICON

Figure A8.1 illustrates the *Deal–Grove model* of thermal oxidation of silicon [1, 2]. The symbols have the following meanings: C_G = Concentration of the oxidant in the bulk of the gas, C_S = oxidant concentration near the oxide surface, C_o = equilibrium concentration of oxidizing species in the oxide at the outer surface, C_i = oxidizing species concentration in the oxide near the oxide–silicon interface, x_{ox} = oxide thickness, F_1 = flux of transport of oxidizing species from the bulk of the gas phase to the gas–oxide interface,

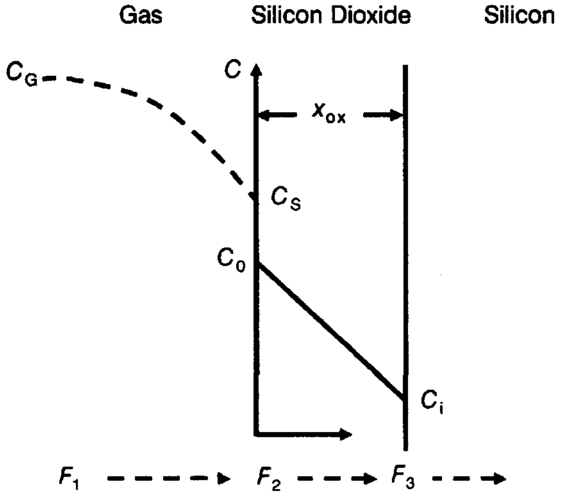


Figure A8.1 Deal–Grove model of thermal oxidation.

F_2 = flux of transport of oxidizing species across the existing oxide toward the silicon, and F_3 = transport flux of the species for reaction at the oxide–silicon interface.

The flux F_1 is proportional to the difference between C_o and C^*, the concentration of the oxidizing species in equilibrium with the oxidant in the gas phase:

$$F_1 = h(C^* - C_o) \tag{A8.1.1}$$

where h is the gas-phase, mass-transfer coefficient.

The flux F_2 can be written as

$$F_2 = D\left(\frac{C_o - C_i}{x_{ox}}\right) \tag{A8.1.2}$$

where D is the diffusivity and $\{(C_o - C_i)/x_{ox}\}$ is the concentration gradient of the oxidizing species across the oxide.

The flux F_3 is expressed as

$$F_3 = k_s C_i \tag{A8.1.3}$$

where k_s is the rate constant of the reaction at the Si–SiO$_2$ interface.

In the steady state we have

$$F_1 = F_2 = F_3 = F \tag{A8.1.4}$$

Substituting for F_1, F_2, and F_3 from Eqs. (A8.1.1), (A8.1.2), and (A8.1.3) into Eq. (A8.14), we get

$$h(C^* - C_o) = D\left(\frac{C_o - C_i}{x_{ox}}\right) = k_s C_i = F \qquad \text{(A8.1.5)}$$

From Eq. (A8.1.5), using the relation $h(C^* - C_o) = F$, we have

$$C_o = \frac{hC^* - F}{h} \qquad \text{(A8.1.6)}$$

Again using the relation $D(C_o - C_i)/x_{ox} = F$ and putting the value of C_o from Eq. (A8.1.6), we get

$$C_i = \frac{hC^* - F}{h} - \frac{Fx_{ox}}{D} \qquad \text{(A8.1.7)}$$

Now applying the relation $k_s C_i = F$ and putting the value of C_i from Eq. (A8.1.7), we obtain

$$F = \frac{k_s C^*}{1 + \dfrac{k_s}{h} + \dfrac{k_s x_{ox}}{D}} \qquad \text{(A8.1.8)}$$

Then oxide growth rate R is given by

$$R = \frac{dx_{ox}}{dt} = \frac{F}{N_{ox}} = \frac{(k_s C^*/N_{ox})}{1 + \dfrac{k_s}{h} + \dfrac{k_s x_{ox}}{D}} \qquad \text{(A8.1.8)}$$

where N_{ox} = density of oxidizing molecules per unit volume of oxide.

Equation (A8.1.18) can be rewritten as

$$dx_{ox} + \frac{k_s}{h} dx_{ox} + \frac{k_s x_{ox} dx_{ox}}{D} = \frac{k_s C^* dt}{N_{ox}} \qquad \text{(A8.1.9)}$$

Integrating Eq. (A8.1.8), we get

$$\int dx_{ox} + \frac{k_s}{h}\int dx_{ox} + \frac{k_s}{D}\int x_{ox}\, dx_{ox} = \frac{k_s C^*}{N_{ox}}\int dt \qquad \text{(A8.1.10)}$$

or

$$x_{ox} + \frac{k_s}{h} x_{ox} + \frac{k_s}{D}\frac{x_{ox}^2}{2} = \frac{k_s C^* t}{N_{ox}} + \text{Constant} \qquad \text{(A8.1.11)}$$

The total oxide thickness x_{ox} consists of two parts: an initial layer of thickness x_{i} that is present on the silicon before oxidation and the oxide layer grown during the step being considered. So, the initial condition is

$$x_{\mathrm{ox}} = x_{\mathrm{i}} \quad \text{at } t = 0 \tag{A8.1.12}$$

Therefore,

$$x_{\mathrm{i}} + \frac{k_{\mathrm{s}}}{h} x_{\mathrm{i}} + \frac{k_{\mathrm{s}}}{D} \frac{x_{\mathrm{i}}^2}{2} = \text{Constant} \tag{A8.1.13}$$

Putting the value of the constant from Eq. (A8.1.13) into Eq. (A8.1.11), we get

$$x_{\mathrm{ox}} + \frac{k_{\mathrm{s}}}{h} x_{\mathrm{ox}} + \frac{k_{\mathrm{s}}}{D} \frac{x_{\mathrm{ox}}^2}{2} = \frac{k_{\mathrm{s}} C^* t}{N_{\mathrm{ox}}} + x_{\mathrm{i}} + \frac{k_{\mathrm{s}}}{h} x_{\mathrm{i}} + \frac{k_{\mathrm{s}}}{D} \frac{x_{\mathrm{i}}^2}{2} \tag{A8.1.14}$$

or

$$x_{\mathrm{ox}}^2 + \frac{k_{\mathrm{s}}}{2D} + x_{\mathrm{ox}}\left(1 + \frac{k_{\mathrm{s}}}{h}\right) = \frac{k_{\mathrm{s}} C^* t}{N_{\mathrm{ox}}} + \frac{k_{\mathrm{s}}}{2D} x_{\mathrm{i}}^2 + x_{\mathrm{i}}\left(1 + \frac{k_{\mathrm{s}}}{h}\right) \tag{A8.1.15}$$

Multiplying both sides by $2D/k_{\mathrm{s}}$,

$$x_{\mathrm{ox}}^2 + 2D\left(\frac{1}{k_{\mathrm{s}}} + \frac{1}{h}\right) x_{\mathrm{ox}} = \frac{2DC^* t}{N_{\mathrm{ox}}} + x_{\mathrm{i}}^2 + \left(\frac{1}{k_{\mathrm{s}}} + \frac{1}{h}\right) x_{\mathrm{i}} \tag{A8.1.16}$$

or

$$\begin{aligned} x_{\mathrm{ox}}^2 + Ax_{\mathrm{ox}} &= Bt + x_{\mathrm{i}}^2 + Ax_{\mathrm{i}} = Bt + \frac{B\left(x_{\mathrm{i}}^2 + Ax_{\mathrm{i}}\right)}{B} \\ &= Bt + B\tau = B(t + \tau) \end{aligned} \tag{A8.1.17}$$

where

$$A = 2D\left(\frac{1}{k_{\mathrm{s}}} + \frac{1}{h}\right), \qquad B = \frac{2DC^*}{N_{\mathrm{ox}}}, \qquad \tau = \frac{x_{\mathrm{i}}^2 + Ax_{\mathrm{i}}}{B} \tag{A8.1.18}$$

τ represents the shift in time coordinate correcting for the presence of initial oxide layer of thickness x_{i}. Equation (A8.1.17) can be expressed as

$$x_{\mathrm{ox}}^2 + Ax_{\mathrm{ox}} - B(t + \tau) = 0 \tag{A8.1.19}$$

This is a quadratic equation in x_{ox} with the solution

$$x_{ox} = \frac{-A \pm \sqrt{A^2 + 4B(t+\tau)}}{2} = \frac{-A \pm \sqrt{A^2\left\{1 + \frac{4B}{A^2}(t+\tau)\right\}}}{2}$$

$$= \frac{-A \pm A\sqrt{1 + \frac{4B}{A^2}(t+\tau)}}{2} = -\frac{A}{2}\left\{1 - \sqrt{1 + \frac{4B}{A^2}(t+\tau)}\right\} \quad \text{(A8.1.20)}$$

where negative value for the square root term has been ignored because it will lead to negative oxide thickness that is inadmissible. Equation (A8.1.20) is recast as

$$x_{ox} = \frac{A}{2}\sqrt{1 + \frac{4B}{A^2}(t+\tau)} - 1 = \frac{A}{2}\sqrt{1 + \frac{t+\tau}{A^2/4B}} - 1 \quad \text{(A8.1.21)}$$

This equation gives the oxide thickness as a function of time. For small oxidation times, $t \ll A^2/(4B)$. Then

$$x_{ox} \cong \frac{B}{A}(t+\tau) \quad \text{(A8.1.22)}$$

where the proportionality factor B/A is called the *linear rate coefficient*. As the oxide growth rate is controlled by the surface reaction rate, the linear rate coefficient depends on the breaking of the bonds at the silicon–silicon dioxide interface and therefore on the crystal orientation.

On the other hand, for long times of oxide growth $t \gg A^2/(4B)$ and $t > \tau$,

$$x_{ox} \cong \frac{A}{2}\sqrt{\frac{t}{A^2/(4B)}} = \frac{A}{2} \times \sqrt{t} \times \frac{2\sqrt{B}}{A} = \sqrt{Bt} \quad \text{(A8.1.23)}$$

This is the well-known parabolic oxidation law. As the parameter B, referred to as the *parabolic rate coefficient*, is governed by diffusion across the oxide already formed on the wafer, it is not related to the crystal orientation.

REFERENCES

1. References 36 of Chapter 8.
2. Reference 37 of Chapter 8.

APPENDIX 8.2: DERIVATION OF EQS. (8.3)-(8.5)

The mathematical basis of diffusion is embodied in the two laws of Fick. The first law states that the rate of transfer of diffusing particles through a unit area is proportional to the concentration gradient measured normal to the area. Hence the flux of atoms in the x direction J_x is given by

$$J_x = -D\frac{\partial N(x,t)}{\partial x} \tag{A8.2.1}$$

where $N(x,t)$ is the concentration of atoms, which depends on both distance and time, and D is a constant of proportionality called the diffusion coefficient. Fick's second law is expressed as

$$\frac{\partial N(x,t)}{\partial t} = D\frac{\partial^2 N(x,t)}{\partial x^2} \tag{A8.2.2}$$

Predeposition

During predeposition, the semiconductor wafer is kept in a furnace at a high temperature surrounded by the vapor of the material to be diffused. At the temperature and vapor pressure of the process, the solute has a solubility in the semiconductor $= N_0$ (say). It is assumed that the vapor pressure is maintained throughout the experiment. Then the diffusion profile achieved in this process is determined by solving Eq. (A8.2.2) subject to the boundary conditions

$$N = 0, \quad x > 0 \quad \text{for } t = 0 \tag{A8.2.3}$$

$$N = N_0, \quad x = 0 \quad \text{for } t > 0 \tag{A8.2.4}$$

Multiplying both sides of Eq. (A8.2.2) by $\exp(-pt)$ and integrating from $t = 0$ to $t = \infty$, we get

$$\int_0^\infty \frac{\partial N}{\partial t}\exp(-pt)\,dt = D\int_0^\infty \frac{\partial^2 N(x,t)}{\partial x^2}\exp(-pt)\,dt \tag{A8.2.5}$$

Integrating the left-hand side of Eq. (A8.2.5) by parts, we obtain

$$\int_0^\infty \frac{\partial N}{\partial t}\exp(-pt)\,dt = \left[N\exp(-pt)\right]_0^\infty + p\int_0^\infty N\exp(-pt)\,dt \tag{A8.2.6}$$

The first term of Eq. (A8.2.6) is $N\exp(-\infty) - N\exp(-0) = 0 - 0 = 0$ because $\exp(-\infty) = 0$ and from Eq. (A8.2.3), $N = 0$ at $t = 0$. Hence Eq. (A8.2.6) reduces to

$$\int_0^\infty \frac{\partial N}{\partial t}\exp(-pt)\,dt = p\int_0^\infty N\exp(-pt)\,dt \tag{A8.2.7}$$

Now the *Laplace transform* of a function $f(t)$ is defined as $\bar{f}(p)$ given by

$$\bar{f}(p) = \int_0^\infty f(t) \exp(-pt)\, dt \tag{A8.2.8}$$

Comparing Eqs. (A8.2.7) and (A8.2.8), the right-hand side of Eq. (A8.2.7) is identified as the Laplace transform of the function N so that Eq. (8.2.7) is rewritten as

$$\int_0^\infty \frac{\partial N}{\partial t} \exp(-pt)\, dt = p\overline{N} \tag{A8.2.9}$$

But the right-hand side of Eq. (8.2.5) is

$$D\int_0^\infty \frac{\partial^2 N(x,t)}{\partial x^2} \exp(-pt)\, dt = D\frac{\partial^2}{\partial x^2}\int_0^\infty N \exp(-pt)\, dt = D\frac{\partial^2 N}{\partial x^2} \tag{A8.2.10}$$

where the order of integration and differentiation has been reversed and the definition of Laplace transform in Eq. (A8.2.8) has been applied.

Combining Eqs. (A8.2.5), (A8.2.9), and (A8.2.10), Eq. (A8.2.5) is modified as

$$p\overline{N} = D\frac{d^2\overline{N}}{dx^2} \tag{A8.2.11}$$

where the partial differentials are full.

Since

$$\overline{N} = \int_0^\infty N \exp(-pt)\, dt \tag{A8.2.12}$$

from the boundary condition (A8.2.4), at $x = 0$

$$\overline{N} = \int_0^\infty N_0 \exp(-pt)\, dt = N_0\int_0^\infty \exp(-pt)\, dt = N_0 \times \left[\frac{\exp(-pt)}{-p}\right]_0^\infty$$

$$= \frac{N_0}{-p}[\exp(-\infty) - \exp(-0)] = \frac{N_0}{-p} \times (0-1) = \frac{N_0}{p}$$

$$\text{that is } \overline{N} = \frac{N_0}{p} \tag{A8.2.13}$$

This means that Eq. (A8.2.11) is to be solved with the boundary condition

(A8.2.13). Rewriting Eq. (A8.2.11) as

$$\frac{d^2\bar{N}}{dx^2} - \frac{p}{D}\bar{N} = 0 \tag{A8.2.14}$$

Comparing Eq. (A8.2.14) with the general linear homogeneous equation with constant coefficients, namely,

$$y'' + a_1 y' + a_2 y = 0 \tag{A8.2.15}$$

it is evident that Eq. (A8.2.14) is a linear homogeneous equation in which $a_1 = 0$ and $a_2 = -p/D$. To solve this equation, we write the characteristic equation

$$m^2 - \frac{p}{D} = 0 \quad \text{or} \quad m^2 = \frac{p}{D}, \qquad \therefore m = \pm\sqrt{\frac{p}{D}} \tag{A8.2.16}$$

So the solution is

$$\bar{N} = A\exp(mx) + B\exp(-mx) \tag{A8.2.17}$$

Since $\bar{N}$ decreases as x increases, the first term is physically inadmissible, giving

$$\bar{N} = B\exp(-mx) = B\exp\left\{\left(-\sqrt{\frac{p}{D}}\right)x\right\} \tag{A8.2.18}$$

At $x = 0$,

$$\bar{N} = B \tag{A8.2.19}$$

But from Eq. (A8.2.13)

$$\bar{N} = \frac{N_0}{p} \tag{A8.2.20}$$

Hence,

$$B = \frac{N_0}{p} \tag{A8.2.21}$$

From Eqs. (A8.2.18) and (A8.2.21), the solution is

$$\bar{N} = \frac{N_0}{p}\exp\left\{\left(-\sqrt{\frac{p}{D}}\right)x\right\} \tag{A8.2.22}$$

Noting that the Laplace transform of the function

$$f(t) = \frac{1}{p}\exp\left\{\left(-\sqrt{\frac{p}{D}}\right)x\right\} \tag{A8.2.23}$$

is

$$\bar{f}(p) = \operatorname{erfc}\left(\frac{x}{2\sqrt{Dt}}\right) \tag{A8.2.24}$$

Eq. (A8.2.22) gives

$$N = N_0 \operatorname{erfc}\left(\frac{x}{2\sqrt{Dt}}\right) \tag{A8.2.25}$$

Drive-In

Here Eq. (A8.2.2) is solved with the initial condition

$$N(x,0) = 0 \tag{A8.2.26}$$

and the boundary conditions

$$N(\infty,t) = 0 \tag{A8.2.27}$$

and

$$\int_0^\infty N(x,t)\,dx = Q \tag{A8.2.28}$$

where Q is the amount of material that has entered the silicon in time t during predeposition.

Let us assume

$$N = \frac{A}{t^{1/2}} \exp\left(-\frac{x^2}{4Dt}\right) \tag{A8.2.29}$$

is a solution of the Eq. (A8.2.2). A is an arbitrary constant. To verify that Eq. (A8.2.29) satisfies Eq. (A8.2.2), the left-hand side of Eq. (A8.2.2) is

$$\frac{dN}{dt} = \frac{d}{dt}\left\{\frac{A}{t^{1/2}} \exp\left(-\frac{x^2}{4Dt}\right)\right\} = A \times -\frac{1}{2} t^{-3/2} \exp\left(-\frac{x^2}{4Dt}\right)$$

$$+A \times \frac{1}{t^{1/2}} \times \exp\left(-\frac{x^2}{4Dt}\right) \times -\frac{x^2}{4D} \times -1 \times t^{-2}$$

$$= A \exp\left(-\frac{x^2}{4Dt}\right)\left\{-\frac{1}{2} t^{-3/2} + \frac{1}{t^{1/2}} \times -1 t^{-2} \times \left(-\frac{x^2}{4D}\right)\right\}$$

$$= A \exp\left(-\frac{x^2}{4Dt}\right)\left(-\frac{1}{2} t^{-3/2} + \frac{x^2}{4D} \times t^{-5/2}\right) \tag{A8.2.30}$$

To calculate the right-hand side, we have

$$\frac{dN}{dx}=\frac{A}{t^{1/2}}\exp\left(-\frac{x^2}{4Dt}\right)\times-\frac{2x}{4Dt} \tag{A8.2.31}$$

Therefore,

$$\begin{aligned}\frac{d^2N}{dx^2}&=\frac{A}{t^{1/2}}\exp\left(-\frac{x^2}{4Dt}\right)\times-\frac{2x}{4Dt}\times-\frac{2x}{4Dt}\\ &\quad+\frac{A}{t^{1/2}}\exp\left(-\frac{x^2}{4Dt}\right)\times-\frac{2}{4Dt}=A\exp\left(-\frac{x^2}{4Dt}\right)\left(\frac{4x^2}{16D^2t^{5/2}}-\frac{1}{2Dt^{3/2}}\right)\\ &=A\exp\left(-\frac{x^2}{4Dt}\right)\left(-\frac{1}{2D}t^{-3/2}+\frac{x^2}{4D^2}t^{-5/2}\right)\end{aligned} \tag{A8.2.32}$$

Therefore,

$$D\frac{d^2N}{dx^2}=A\exp\left(-\frac{x^2}{4Dt}\right)\left(-\frac{1}{2}t^{-3/2}+\frac{x^2}{4D}t^{-5/2}\right) \tag{A8.2.33}$$

Comparing the left and right sides of Eq. (A8.2.2), given by Eqs. (A8.2.30) and (A8.2.33), we find that Eq. (A8.2.29) is a solution of Eq. (A8.2.2). It is further noted that Eq. (A8.2.29) satisfies the boundary conditions of the experiment, namely, (i) it is symmetrical with respect to $x=0$. (ii) It approaches zero as $x\to\pm\infty$ for $t>0$. (iii) For $t=0$, it vanishes everywhere except at $x=0$ where it is infinite. Now the quantity of dopant diffusing into the semiconductor is

$$Q=\frac{1}{2}\int_{-\infty}^{\infty}\frac{A}{t^{1/2}}\exp\left(-\frac{x^2}{4Dt}\right)dx \tag{A8.2.34}$$

Let us change the variable to $u^2=x^2/(4Dt)$. Then we get

$$Q=\frac{1}{2}\int_{-\infty}^{\infty}2AD^{1/2}\exp(-u^2)\,du=2AD^{1/2}\pi^{1/2} \tag{A8.2.35}$$

From Eqs. (A8.2.29) and (A8.2.35), we have

$$N=\frac{Q}{(\pi Dt)^{1/2}}\exp\left(-\frac{x^2}{4Dt}\right) \tag{A8.2.36}$$

The expression for Q given by Eq. (8.5) is obtained by integrating $N(x,t)$ over x. From Eqs. (A8.2.28) and (A8.2.25), we have

$$Q = \int_0^\infty N(x,t)\,dx = \int_0^\infty N_0 \mathrm{erfc}\left(\frac{x}{2\sqrt{Dt}}\right) dx$$

$$= N_0 \int_0^\infty \mathrm{erfc}\left(\frac{x}{2\sqrt{Dt}}\right) dx = N_0 \frac{1/\sqrt{\pi}}{1/(2\sqrt{Dt})} = 2N_0\sqrt{\frac{Dt}{\pi}} \qquad \text{(A8.2.37)}$$

where the equation

$$\int_0^\infty \mathrm{erfc}(x)\,dx = \frac{1}{\sqrt{\pi}} \qquad \text{(A8.2.38)}$$

has been used.

9

POWER IGBT MODULES

9.1 PARALLELING IGBTs, AND INTEGRATION OF LOGIC CIRCUITS WITH POWER COMPONENTS

In the preceding chapters, we have studied IGBT as a single discrete component encapsulated in a commercial package like TO-218, TO-3, TO-220, and so on, for low-power applications with current ratings up to 100 A. Unlike high-power diodes and thyristors (wherein a single large-area die is housed in pressure contacted hermetically sealed hockey puck packages), medium- and high-power IGBTs comprise (a) several paralleled chips assembled in a plastic cover, (b) a non-airproof package containing injected silicone gel, and (c) Al wire bonded connections between the main emitter electrode and the IGBT chips. These chips are generally cooled from the collector side only. This paralleled construction is favored because an IGBT device consists of millions of elementary cells connected in parallel, and manufacturing large-area IGBTs suffers from poor processing yield, principally due to gate oxide defects, limiting the chip size to 2–3 cm^2. Obviously, therefore, the power density in the IGBT is restricted by physical reasons. Thus several chips must be paralleled to increase the current-carrying capability. Hence, the concept of IGBT modules originates from paralleling several IGBT chips together, to create a single packaged high-power device [1–10]. It must be reiterated that at the device level the larger the chip size of the IGBT, the poorer the yield. However, the interconnection of a large

---

The Insulated Gate Bipolar Transistor (IGBT) Theory and Design, By Vinod Kumar Khanna.
ISBN 0-471-23845-7 © 2003 Institute of Electrical and Electronics Engineers

number of IGBTs in parallel opens up new perspectives and applications for IGBTs.

In power electronic circuits such as rectifier bridges, converters, and so on, all power semiconductor devices including the free-wheeling diodes are assembled on one substrate, producing a "power module." Integration of more microelectronic functions is performed by adding simple logic circuits —for example, gate drive, self-protection circuits, and microprocessor capability, leading to "smart power module." As the adjective "smart" implies, "smart power" is a family of devices/systems which combines intelligence, or brain power with force, or muscles, on the same substrate or in one chip (smart power IC). "Smart" imparts versatility and flexibility to the system, whereby its characteristics and parameters are settable via software modification.

While the main objective of a power module is to increase the current capacity, it also provides increased functionality by including, typically, (i) an *input single phase bridge rectifier* with four rectifier diodes, (ii) an *inverter stage* consisting of six IGBTs with their free-wheeling diodes, (iii) a *gate driving circuit*, and (iv) *protection circuits* for overcurrent, short circuit, overtemperature, and undervoltage lock-out (due to control supply failure). Fault output signals are generated for external monitoring of protection circuits.

The basic philosophy of a power module is to reduce the infrastructural requirement for a given application, thus leading to overall cost-effectiveness. The module streamlines the construction and assembly for many power systems, leading to savings in components, mounting, assembly, packaging and external wiring costs. Thus power modules have dramatically reduced system costs, while their software adaptation to particular requirements has reduced engineering and manufacturing costs by reducing the number of components required for higher-order systems. (However, this is true only on a case-by-case basis and varies as a function of power levels, design time, and production volume. For example, the cost of a TO-220 package for discrete devices together with ease of cooling, is very attractive). Other benefits include (a) increased system reliability due to fewer components and solder joints and (b) built-in self-overcurrent and over-temperature protection capabilities, which minimize failure rate.

The smart power market is a major segment of the semiconductor market, offering design simplification by combining the functions of several discrete devices and ICs into one chip. Applications of smart power ICs include motor controls, especially for the small DC stepper or synchronous motors, used in computer peripheral equipment, robotics and automotive electronics, electromechanical control for printers, switching regulators, audio amplifiers, and so on.

Example 9.1 Given a DC supply of 150 V and load resistance of 1 Ω, and two IGBTs of current ratings 100 A, what resistance should be connected in series with each IGBT to ensure a reasonable current sharing?

Figure E9.1.1 shows the two IGBTs of the same specifications, connected in parallel. Both the IGBTs have the same voltage drop across them when conducting.

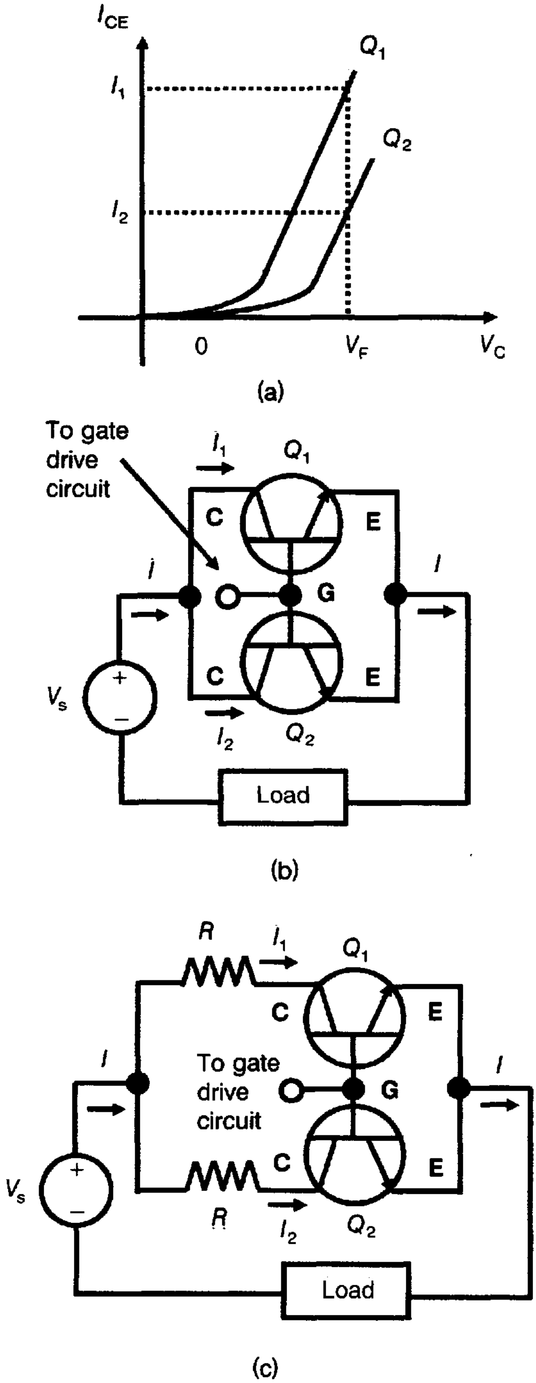

I_{CE}
Q_1
I_1
Q_2
I_2
0
V_F
V_C
(a)
To gate drive circuit
I_1
Q_1
C
E
I
G
I
V_s
+
−
C
E
I_2
Q_2
Load
(b)
R
I_1
Q_1
C
E
I
To gate drive circuit
G
I
V_s
+
−
C
E
R
I_2
Q_2
Load
(c)

Figure E9.1.1 Current sharing by IGBTs. (a) Forward characteristics of IGBTs Q_1 and Q_2. (b) Parallel-connected IGBTs Q_1 and Q_2. (c) Current sharing in paralleled IGBTs Q_1 and Q_2 with resistors R.

However, because of differences in the ON-state characteristics of the two IGBTs (even if very small), the currents flowing through them will be different. Series-connected resistors are used to swamp the inequality of the IGBT potential drops. Let the current flowing through IGBT Q_1 be I_1 and let that through IGBT Q_2 be I_2 ($I_1 > I_2$). Then $I = I_1 + I_2$. Now the maximum load current = 150/1 = 150 A. From safety considerations, $I_1 \cong I_2$. Generally, it is aimed that the difference in currents be ≤ 20%. Then, if $I_1 = 90$ A, we have $I_2 = 60$ A. Because I_1 is higher than I_2, it is desired that Q_1 be operated at lower forward voltage than Q_2 for equitable current sharing. Let V_1 and V_2 be the minimum and maximum permissible potential differences across Q_1 and Q_2, respectively. Then applying Kirchhoff's voltage law, we get

$$RI_1 + V_1 = RI_2 + V_2 \tag{E9.1.1}$$

This gives

$$R = \frac{V_2 - V_1}{I_1 - I_2} \tag{E9.1.2}$$

Choosing arbitrarily the minimum and maximum allowed voltage values across the two IGBTs, namely, $V_1 = 3.45$ V and $V_2 = 3.55$ V, we find that $R = (3.55 - 3.45)/(90 - 60) = 3.33$ mΩ. The voltage drops across the two resistors are $RI_1 = 3.33 \times 10^{-3} \times 90 = 0.2997$ V $\cong 0.3$ V and $RI_2 = 3.33 \times 10^{-3} \times 60 = 0.1998$ V $\cong 0.2$ V. Thus there is an extra power loss due to the voltage drop = 0.3 + 0.2 V = 0.5 V across the resistors, but this is less than 7.14% of the power loss due to voltage drop across the IGBTs = 3.45 + 3.55 = 7 V.

Note: The series resistors equalize the currents through parallel IGBTs but do not compensate for unequal turn-on times or latching currents. For example, in a parallel-connected set of IGBTs, if one IGBT turns on before the remaining IGBTs, it momentarily carries the full load current. This is not serious unless the surge capability of the IGBT is not exceeded. For transient current sharing, inductors are used in place of resistors.

Example 9.2 A 2-kV DC source supplies power to a load. The power conditioning is carried out by IGBTs. The available IGBTs are rated at 500 V, 20 A. How will you plan to execute the steady-state voltage sharing?

Refer to Fig. E9.2.1. Let the breakdown voltage of each IGBT be V_B. As the two IGBTs together block the supply voltage V_s, if V_1 and V_2 are the potential drops across these IGBTs, then $V_s = V_1 + V_2$. To prevent disproportionate voltage sharing among a string of IGBTs, resistors R are connected in parallel with each IGBT. In order that the forced sharing of voltage be effective, the current flowing through the resistors must be large enough to swamp the inequality in the IGBT leakage currents. Let there be N IGBTs. For worst-case calculations, it is assumed that the value of R is the maximum value for current limiting. Furthermore, the leakage current of IGBT Q_1 is assumed to be negligible, while that of other IGBTs is assumed to be a maximum value I_m. It is also assumed that the supply voltage V_s is the maximum that can be applied when resistor R is used, and the potential drop V_1 across IGBT Q_1 is the peak voltage rating of the IGBT, the maximum possible value. Applying Kirchhoff's voltage law, we get

$$V_s - V_1 = (N - 1)V_2 \tag{E9.2.1}$$

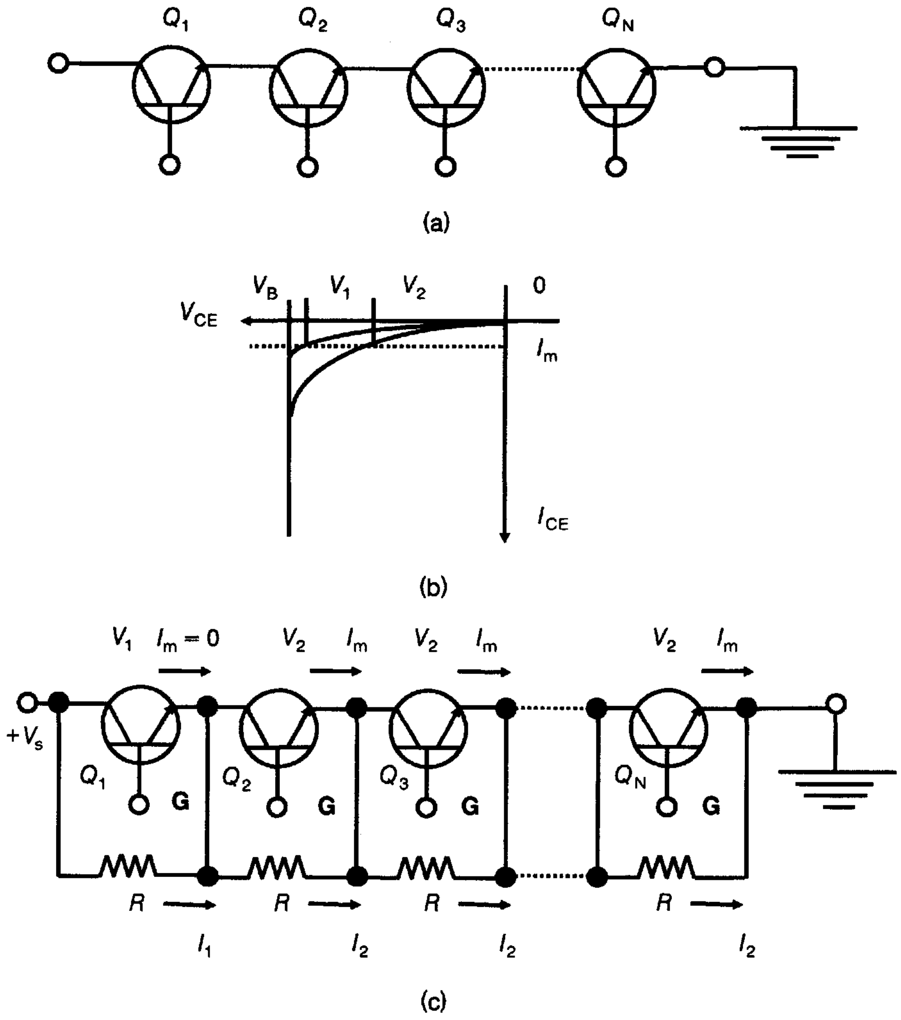

Q_1
Q_2
Q_3
Q_N
(a)
V_B
V_1
V_2
0
V_{CE}
I_m
I_{CE}
(b)
V_1
$I_m = 0$
V_2
I_m
$+V_s$
Q_1
Q_2
Q_3
Q_N
G
R
I_1
I_2
(c)

Figure E9.2.1 Voltage sharing by IGBTs. (a) IGBTs Q_1 and Q_2 in series. (b) Blocking characteristics of IGBTs Q_1, $Q_2 = Q_3 = \cdots Q_N$. (c) Steady-state voltage sharing of IGBTs with resistive voltage divider.

From Kirchhoff's current law, we have

$$I_1 = I_m + I_2 \tag{E9.2.2}$$

By Ohm's law,

$$I_1 = V_1/R \quad \text{and} \quad I_2 = V_2/R \tag{E9.2.3}$$

These equations yield

$$R = \frac{NV_1 - V_s}{(N-1)I_m} \tag{E9.2.4}$$

Now, four IGBTs in series will not block 2000 V due to the differences in their characteristics. But the use of too many IGBTs will unnecessarily increase the cost. Let us choose $N = 5$ and the high leakage current value $I_m = 1$ mA. Then, $R = \{5 \times 500 - 2000\}/\{(5-1) \times 1 \times 10^{-3}\} = 125$ kΩ. Since, $V_s - V_1 = (N-1)V_2$, we have $V_2 = (V_s - V_1)/(N-1) = (2000 - 500)/4 = 375$ V. Hence $I_2 = V_2/R = 375/(125 \times 10^3) = 3$ mA. Thus the currents flowing through the resistors must be three times the IGBT leakage currents for swamping the leakage current effects.

Note 1: When the IGBTs are turned on or off, the sharing resistors become ineffective in ensuring equal voltages across them. IGBTs which recover their blocking state fastest or which turn on last, will be exposed to the full voltage of a series string, often disastrously. The transient voltage distribution varies inversely with the device capacitance, and inequality is prevented by connecting a capacitor in parallel with each IGBT, having a value greater than the device capacitance to slow down the rate of voltage change across the IGBT during turn-on and turn-off. A small resistor is also connected in series with the capacitor to limit the discharge current of the capacitor, which would otherwise occur when the IGBT is turned on.

Note 2: Voltage balancing by above method provides simplicity and robustness but increases the power losses and commutation times. Smaller energy losses and higher working frequencies are obtained by snubberless gate side techniques.

9.2 POWER MODULE TECHNOLOGIES

Depending on the power ratings of devices, methods differ with regard to the substrate material of choice, the copper deposition technique for producing current leads on the substrate, and chip mounting and interconnection procedures [11].

9.2.1 Substrates and Copper Deposition

Two types of substrates are commonly employed for modules: *insulating substrates* and *insulated metal substrates*. Amongst the *insulating substrates*, the varieties generally used are aluminum oxide (alumina, 96%) and aluminum nitride (AlN). Beryllium oxide (beryllia, BeO) is sometimes preferred due to its high thermal conductivity, but is toxic and costly. Product safety requirements may preclude the use of beryllia due to toxicity. These substrates are coated with a Cu layer by one of the following methods: *copper thick film* (CTF) *technology* using screen printing, *copper on ceramic* (COC) by electroless plating, and *direct copper bonding* (DCB) through eutectic oxide bonding. The sheet resistance of a single coating of Cu of thickness 28 μm is 1.1 mΩ/□ while that of a Cu sheet (having thickness = 300 μm) in DCB substrates is 0.07 mΩ/□. The former is used for low current (< 20 A), while the latter can carry currents of more than 200 A.

In the CTF technique (Fig. 9.1a), the substrate coated with a conducting ink like silver-palladium is fixed to the 3-mm-thick Cu baseplate below by a

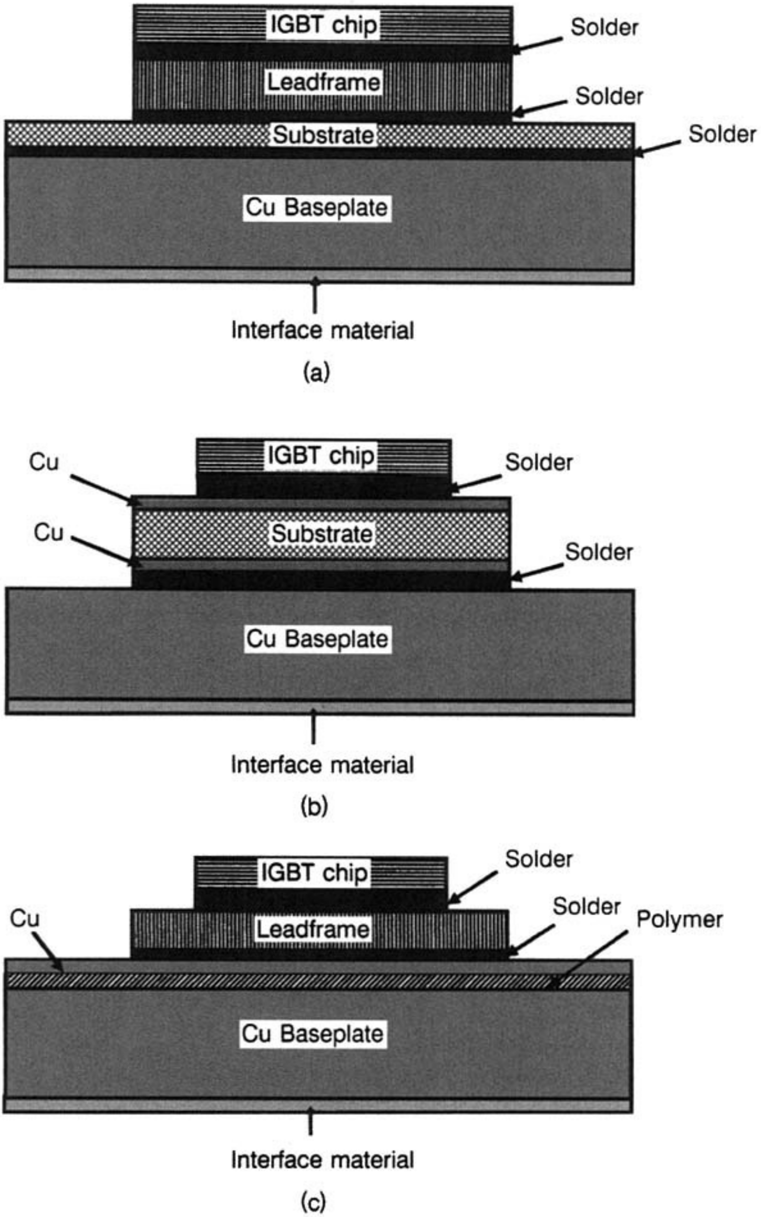

IGBT chip
Solder
Leadframe
Solder
Substrate
Solder
Cu Baseplate
Interface material
(a)
Cu
IGBT chip
Solder
Cu
Substrate
Solder
Cu Baseplate
Interface material
(b)
IGBT chip
Solder
Cu
Solder
Polymer
Leadframe
Cu Baseplate
Interface material
(c)

Figure 9.1 Layered structure of three common stacking technologies for power modules. (a) Thick-film hybrid stack construction. (b) Direct-bonded copper stack construction. (c) Insulated metal substrate assembly.

solder layer, and it is fixed to a leadframe (thickness = 0.9 mm) above by another solder layer. The IGBT chip is soldered on the leadframe. The leadframe is necessary because the thick film cannot carry high currents. It also aids in spreading heat. However, its use requires additional solder layers. CTF is especially attractive when fabricating hybrid modules containing integrated circuits; and components, like resistors, can be fabricated with thick film techniques.

In the DCB method (Fig. 9.1b), a 0.3-mm-thick Cu sheet is bonded on both the sides to ceramic substrate (0.6 mm thick). This is done by growing a thin oxide on Cu sheet and placing the Cu sheets and the substrate in intimate contact at high temperature. During this high-temperature treatment, the thin oxide layer on Cu is chemically bonded with the substrate. Subsequently, the Cu–substrate–Cu sandwich is cooled down to room temperature when additional strength is imparted to the bonding. The substrate is subjected to compression due to the high thermal coefficient of expansion of Cu. Apart from the absence of the solder layer as in thick film method, the high dielectric strength of the ceramic substrate and its high thermal conductivity have made the DCB the popular stack construction scheme.

The insulating ceramic substrates pose a restriction on the size of DCB modules due to the large mechanical stress applied during bonding to copper layers and covering by some mould in the housing. The size limitation is overcome by metal substrates. So, for small modules, the insulating substrates are superior, whereas for the larger modules, the metal-based substrates are useful. In the insulated metal substrates, an insulating layer of adequate dielectric strength and thermal conductivity is deposited producing the types of substrates as follows: (i) Porcelain Enameled metal substrates (PEMS) made of either low-carbon steel or Cu-clad invar coated with enamel. (ii) Metal substrates with plasma spraying of alumina films. These are covered with high-velocity powder-sprayed Cu layers, patterned during deposition, using laser-cut copper masks. (iii) Polymer insulated metal substrates (PIMS). These comprise Al, steel, or molybdenum plates covered with a thin coating of epoxy or polyimide (50–100 μm), then laminated with a sheet of Cu over which the leadframe (thickness = 1.25 mm) is attached (Fig. 9.1c). The leadframe also acts as a heat-spreader, improving the thermal performance of the module. Among the shortcomings of this method are (a) the low thermal conductivity of polymer, raising the thermal resistance of the stack, and (b) capacitive cross-talk problems associated with the thin polymer layer.

It is worthwhile noting that low-power modules, up to 50 A, 600 V and 15 A, 1200 V, use a multilayer epoxy-based isolation system in which alternating layers of Cu and epoxy produce a shielded printed circuit on an Al base plate. In the medium-power range, up to 75 A, 600 V and 25 A, 1200 V, DCB ceramic substrates are used with the IGBTs mounted on DCB substrate and accessory circuits on a separate PCB. At high-power levels, up to 200 A, 600 V and 100 A, 1200 V, aluminum nitride substrates are employed.

The case-to-sink interface has a great impact on the heat flow in power modules, accounting for ~ 50% of the thermal resistance. Interface materials are chosen from electrical, thermal, thickness, and cost considerations. They are characterized in terms of their thermal conductivity, volume resistivity, dielectric strength, operating temperature range, stability, long-term reliability, nontoxicity, shelf life, and so on. They are divided into the following generic classes [12]:

(i) *Thermal Greases* These are silicone-free or silicone-based greases or paraffin-based compounds, formulated to enhance heat transfer across the interface. Typically, their thermal conductivity is 0.7 W/m-°C, dielectric strength is 200–300 V/mil, and volume resistivity is 10^{12}–10^{15} Ω-cm. Their operating temperature range is −40°C to 200°C. They do not dry out, melt, or harden upon continuous long exposures to high temperatures (200°C). The highly lubricating base oil used in some types effectively fills the microscopic air gaps between mating surfaces, exhibiting low interface thermal resistance ~ 0.02°C-in.^{2}/W. Their shelf life is indefinite, if unopened.

Doping of the thermal grease with Cu powder decreases the thermal resistance of the grease appreciably, but the grease is thickened and becomes difficult to handle. Good solder joints have a small thermal resistance but are prone to fatigue, requiring repair and replacement.

(ii) *Thermal Adhesives* Epoxy adhesives are ideal for large-scale applications, providing efficient heat transmission with high-voltage isolation capability, besides low shrinkage, coefficients of thermal expansion compatible to Cu or Al, and good adhesion. Their thermal conductivity is slightly better than that of greases: ~ 0.8–1.3 W/m-°C. Dielectric strength is also superior: ~ 500–1500 V/mil. But operating temperature range is toward the lower temperature side—for example, from −65°C to 155°C. Also, shelf life is limited, around a year.

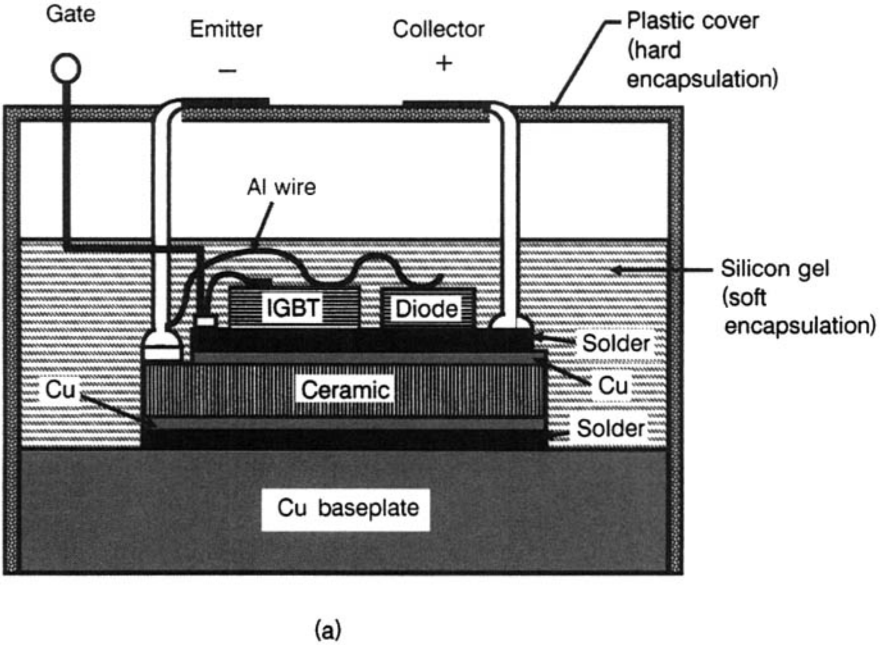
(iii) *Thin Films* These are cost-effective alternatives to thermally conductive greases. They are applied with commercial hot stamping equipment. Their thermal conductivity is ~ 3 W/m-°C, their volume resistivity is 10^{-6} to 10 Ω-cm, and their thickness is around 0.1 mm.

(iv) *Insulating Films* These are low-cost polyimide plastic films serving as a substitute for mica. They have a high resistance to flow when compressed, particularly at high temperatures. Their mechanical and electrical properties are maintained over a broad range of temperatures and frequencies.

(v) *Double-Sided Thermal Tapes* These are easy to apply and need no curing.

(vi) *Thermal Interface Pads* These pads, using nylon-base resin or dialyl phthalate, are thicker than tapes and are either of conducting or isolating type.

(vii) *Gap Fillers* These are soft, flexible, elastic materials to fill gaps between hot components, allowing them to blanket uneven surfaces.


Gate
Emitter
–
Collector
+
Plastic cover
(hard
encapsulation)
Al wire
IGBT
Diode
Silicon gel
(soft
encapsulation)
Solder
Cu
Ceramic
Cu
Solder
Cu baseplate
(a)

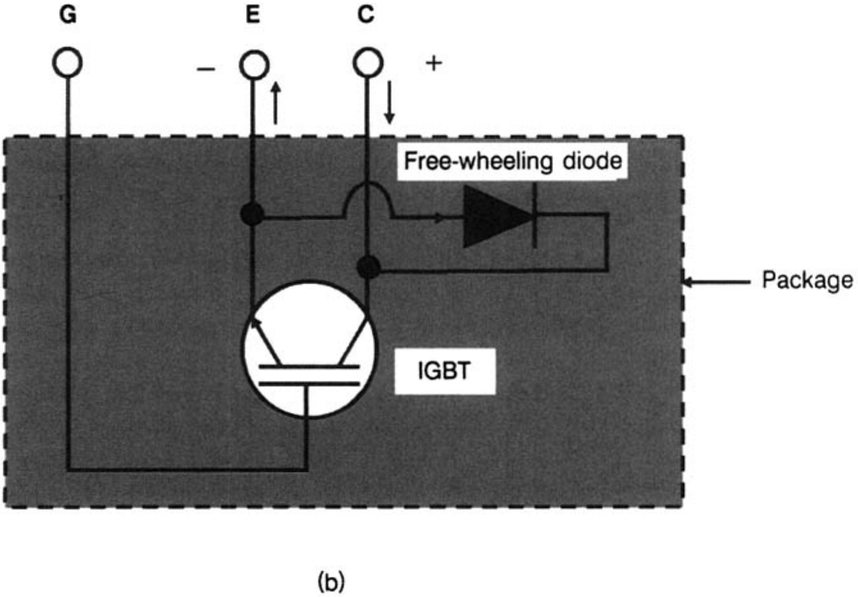

G
E
C
–
+
Free-wheeling diode
Package
IGBT
(b)

Figure 9.2 (a) Power module containing the IGBT chip and the free-wheeling diode. (b) Wiring diagram of the module.

9.2.2 Chip Mounting

Bare IGBT dice are normally supplied with a solderable metallization like Ti–Ni–Ag/Au on the back side of the chip and wire bondable Al metallization on the front. Therefore, the back plane of the die is soldered on the proper landing area of the module substrate by fluxless high melting point (310°C) Pb (95%)–Sn (5%) solder. The solder preforms are carefully placed in the milled holes of a graphite cast along with various dice and clips. Passage through a hydrogen-flushed conveyer belt furnace completes the soldering. Alternatively, a eutectic die bonder is used as a pick-and-place machine. The use of soft solder (lead–tin alloys) is also widespread. Soft solders yield under stress by plastic deformation and thereby protect the thin silicon chips from rupturing. The low melting point of soft solder affords the advantage of easy construction of modules, without any deterioration of the chip behavior. But they suffer from fatigue under cyclic plastic deformation. This deformation occurs as the IGBT chip is heated up and cooled down by intermittent electrical power. After several thousand cycles, the chip begins to malfunction. This is a serious concern. Reliability of IGBT modules will be discussed in Section 9.8. In low-power hybrid technology, Ag-filled adhesives, applied by dispenser and screen printing, may alleviate this concern.

9.2.3 Interconnections and Packaging

Heavy-wire wedge–wedge bonds are used for chip to substrate and chip-to-chip interconnects. A single Al wire (175-μm diameter) fails at 4.9 A, depending on duty cycle and ambient temperature. To increase the current-carrying capacity, the wires are paralleled severalfold. Wire diameters ranging from 300 to 500 μm are used. Wire bonding is fully automated, making it cost effective.

Plastic transfer molding is used for packaging. The wire-bonded assemblies are clamped and placed in a thermosetting resin. They are transferred to a mold, which has cavities for several packages. The encapsulation material is preheated by RF induction heaters and injected into the feed vessel of the mold from where it enters the cavities. After pressuring and curing, the mold is opened and the packages are removed. Figure 9.2 shows the construction of an IGBT module with a free-wheeling diode.

9.3 ISOLATION TECHNIQUES

Now we move on from the PCB or hybrid modular packaging to the monolithic or smart power ICs. Like the module technologies, smart power IC technologies may be classified in terms of isolation technique [13] as (i) dielectric isolation, (ii) self-isolation, and (iii) junction isolation.

9.3.1 Dielectric Isolation (DI)

Here the chip contains monocrystalline Si islands in a polycrystalline substrate, with the mono- and poly-Si separated by an intervening layer of SiO_2 (Fig. 9.3a). To produce the starting material, N^+ layers are locally diffused in a single-crystal Si wafer. This is followed by patterning of the grooves, their anisotropic etching, and their oxidation. Then a thick polySi layer is deposited. Finally, wafers are lapped and polished from the single-crystal side to delineate the islands for device fabrication. Another promising DI technique relies on silicon-on-insulator (SOI) substrate fabricated by silicon direct bonding with isolation provided by narrow trenches filled with silicon dioxide and polysilicon.

Dielectric isolation is considered as a complete isolation because the low-voltage circuitry is not the least affected by high-voltage sections. The additional attraction is that at high temperatures the leakage current of PN junctions increases to intolerable magnitudes. Here, the only coupling between the components is through capacitance. Besides the reduced leakage current, other advantages include low parasitic capacitance of components to substrate, freedom from latchup and crosstalk between elements, and smaller die size for high-voltage components. The main disadvantage of dielectric isolation is the exorbitant cost of starting material mainly incurred in the processing operations for fabricating the Si islands. Other drawback of dielectric isolation is the lower thermal conductivity of the oxide layer and the insulating polycrystalline substrate, as compared to monocrystalline Si, thereby decreasing the efficiency of heat dissipation and restricting its use in high-power applications.

Dielectric isolation between power and intelligence, provided by silicon-on-insulator (SOI) technology, along with easier design possibilities and elimination of parasitics, will enable monolithic integration of several IGBTs in bridge configuration. However, its present cost is prohibitively high.

9.3.2 Self-Isolation

Self-isolation is realized by making a PN junction between the diffused or implanted layer and the substrate (Fig. 9.3b). Upon reverse-biasing of the junction, its resistance increases. In this way, neighboring junctions are disassociated. For example, in a MOSFET, the source and drain PN junctions isolate themselves during reverse biasing. This method is mainly used in MOS processing line, especially in VLSI, when the integrable devices do not require complete isolation due to sharing of some terminals.

The process is simple, consisting of CMOS or DMOS fabrication sequences with some extra steps for additional structures. It does not need any deep diffusion. The P-type Si wafer is coated with a thin layer of SiO_2 followed by deposition of Si_3N_4. After patterning and etching the nitride

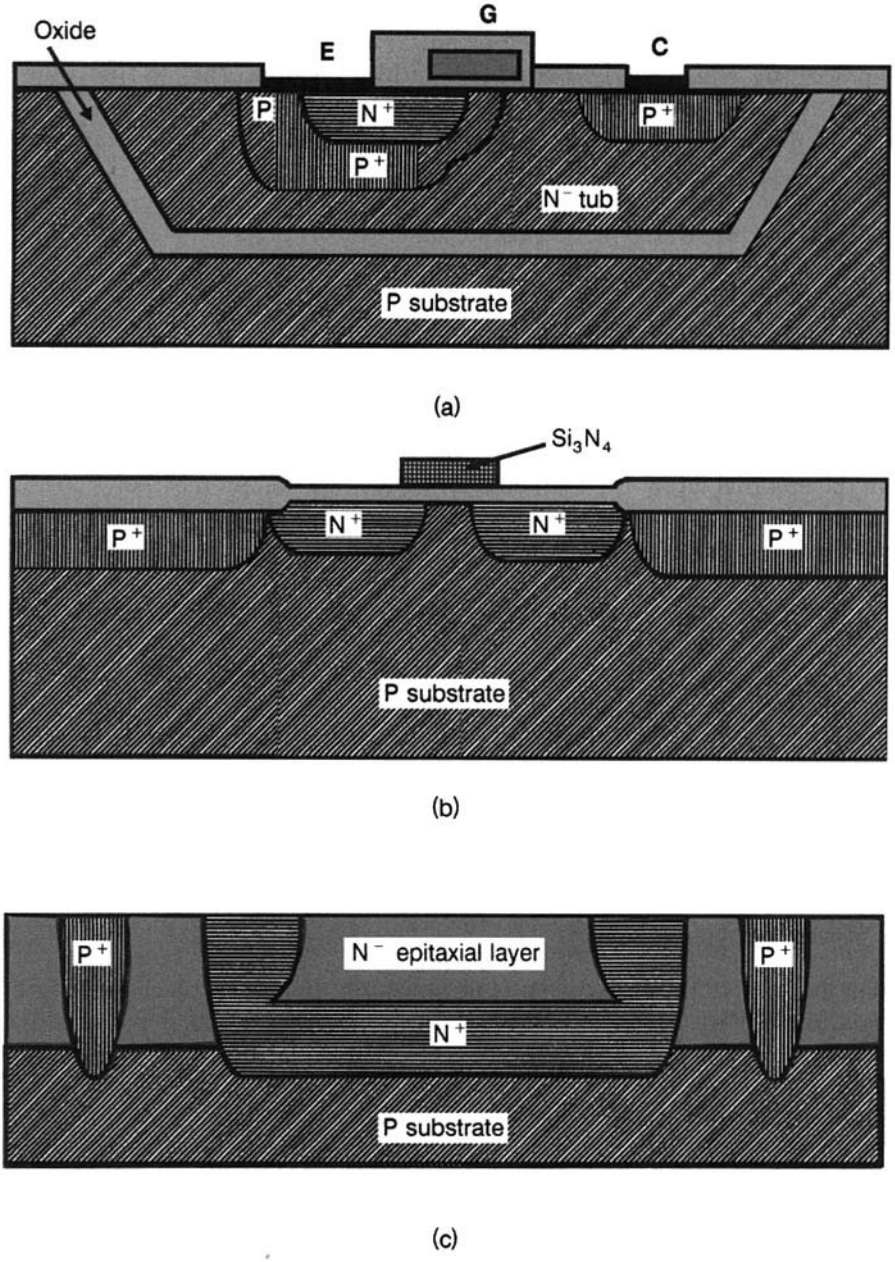

Oxide
E
G
C
P
N +
P +
P +
N⁻ tub
P substrate
(a)
Si_3N_4
P +
N +
N +
P +
P substrate
(b)
P +
N ⁻ epitaxial layer
P +
N +
P substrate
(c)

Figure 9.3 Schematic representation of isolation schemes. (a) Dielectric isolation. (b) Self-isolation. (c) P–N junction isolation.

layer, P^+ regions are formed by ion implantation. A thick layer of SiO_2 is grown over the areas unprotected by nitride. Then the nitride film is etched, poly-Si gate is formed, and N-type impurities are implanted in source and drain regions. The PN junctions thus created are self-isolating.

But the process suffers from lack of flexibility and susceptibility to parasitic effects related to biasing and supply voltages. However, high-voltage and high-current components are easily integrated, because the process involves the power DMOSFET fabrication steps.

9.3.3 PN Junction Isolation

Here the N-type epitaxial layer is surrounded by P-type Si (Fig. 9.3c). So, reverse-biased PN junctions separate the Si islands. The diffusion depth of impurities to join the P^+ region with the P^- substrate well restricts the depth of the epitaxial layer. The method offering the best compromise between cost and versatility is extensively utilized in bipolar integrated circuits. Voltages up to 1200 V are achieved with lateral DMOSFETs by applying the REduced SURface Field (RESURF) technique.

Starting with a P^- substrate, photolithography is used to make the buried layer at selected sites, through As or Sb implantation. Then an N^- epitaxial layer of desired thickness is grown, followed by oxidation, patterning, and boron and phosphorus implantations. The end result is the formation of P^+ regions separating N^- islands in a P^- substrate. Through the N^+ diffused regions, the buried layer is accessed.

9.4 INTEGRABLE DEVICES: BIPOLAR, CMOS, DMOS (BCD), AND IGBT

The term "BCD" refers to the family of mixed processes, allowing integration of bipolar, CMOS, and DMOS transistors onto a single chip; the IGBT is another arrival in this family [13]. The architecture of the first-generation BCD process was based on the polySi gate, self-aligned, VDMOS technology. Bipolar and CMOS components were added by applying the same dopant profiles and polySi layer in a standard junction isolation process. The minimum feature size in the lithography was 4 μm, and voltage limit was up to 250 V. The core of the second-generation BCD was CMOS for reducing circuit design complexity. In the third-generation BCD, at 1.2 μm, nonvolatile memories were integrated, making power ICs system-oriented. BCD is a quasi-vertical technology, seeking the advantages of both the lateral and vertical disciplines. The fifth-generation BCD, designed at 0.6 μm, includes EEPROM.

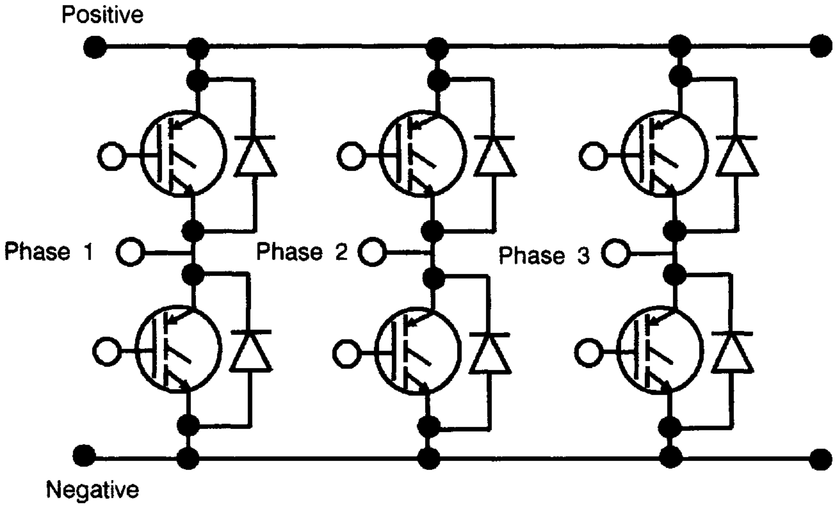


Figure 9.4 Three-phase bridge configuration.

9.5 POWER IGBT DRIVING, TEMPERATURE SENSING, AND PROTECTION

High-speed (20 kHz), intelligent power modules (IPMs) in the 0.4- to 22-kW range, containing six or seven high-performance IGBTs (Fig. 9.4) having current ratings up to 200 A at 600 V and 100 A at 1200 V, in a three-phase bridge configuration (plus the seventh IGBT for dynamic braking), along with integrated gate drive and protection circuits (overcurrent, short-circuit, overtemperature, and control supply failure), are used for motor drives [14]. The IGBT chips have channel lengths of 1.0–0.5 μm. They differ from the conventional chips in two respects. First, they have an auxiliary current sense emitter, which is used by the short-circuit and overcurrent protection circuits of the module for safeguarding the chips. Second, the SOA capability of the IGBT chips matches the safety limits provided by the internal control circuits of the module. In hard-switching applications, turn-on losses of the free-wheeling diodes associated with each IGBT are due to larger recovery current and longer recovery time. An ultrafast and soft recovery diode is fabricated by reducing the doping concentration and depth of anode-side P layer, with insignificant increase of forward drop, thereby reducing turn-on losses.

The gate drive circuit is optimized to minimize switching losses. The peak gate drive current is adjusted to obtain fast switching with due consideration to surge voltages, *dv*/*dt*, and so on. Protection circuits safeguard the IGBTs from damage during system malfunctioning or overstress. From the current

Table 9.1 Classification of Faults in a Converter

Serial No.	Type of Fault	Reason	Result
1	Current faults	**Overcurrent:** Reasons include control error and load impedance reduction.	Slow rise of collector current and destruction.
		Short-circuit current: Possible reasons are control error, destruction of IGBT, diode, or load overcurrent.	Rapid rise of collector current and destruction.
2	Over-temperature	Occurs when the maximum permissible temperature is exceeded due to high-power losses or case temperatures.	Destruction of the device.
3	Overvoltage	Occurs due to dynamic voltage drop in parasitic inductance during IGBT turn-off.	IGBT destruction if the breakdown voltage is exceeded.

sense emitter of the IGBT chips, the sense current is diverted through a resistor network, which provides the control signal for the gate control IC. This IC has comparator circuits to detect short-circuit or overcurrent conditions. High di/dt conditions are avoided by controlled dv/dt shutdown methods. A real-time control circuit, clamps the short-circuit current to a safe limit, thereby providing sufficient time for the control circuit to initiate shutdown. With the help of a temperature sensor, attached to the isolated base plate near the IGBT, the power devices are turned off together by fault signal generation, when the base plate temperature exceeds a predecided level. If the supply voltage falls below the specified value (15 V) for control circuits, the IGBTs are turned off and a fault signal is produced. Automatic resumption of the operation takes place when the supply voltage reverts to the correct value. Table 9.1 lists the common faults occurring in converters, and Table 9.2 summarizes the different protection schemes in IGBT modules.

Merely turning off the IGBT in the event of a fault condition does not suffice. This may itself damage the device if proper precautions in reducing the voltage transients are not taken. Therefore the detrimental switching voltage transients causing overvoltage stresses must be understood. Upon abruptly turning off the IGBT, the energy trapped in the circuit stray inductance, also known as the "DC loop inductance," results in a voltage overshoot across the device. Two factors determine the amount of voltage

Table 9.2 Built-in Protection Schemes in IGBT Power Modules

Serial No.	Protection Mechanism	Description	Indication
1	Overcurrent protection	This switches off the individual IGBT when the current in the same inverter leg exceeds a certain magnitude. However, the operation is delayed by ~5 μsec to avoid unnecessary tripping during the recovery of the associated free-wheeling diode.	The indication is effected by a logic output from the module. The "fault-out" signal has duration of 1 msec. During this period, pulses from the PWM input are inhibited. The module resets to "normal operation" with the arrival of a new pulse after 1 msec.
2	Short-circuit protection	The current magnitude for short-circuit tripping is set at 1.5 times the overcurrent trip level. Protection is realized by a two-stage shutdown when the gate–emitter voltage is decreased immediately to half the first stage voltage. Voltage transient generated during the turn-off is thus limited.	As above.
3	Over-temperature protection	This is accomplished through a temperature sensor mounted on the base substrate of the module. It therefore provides the steady-state module temperature rather than instantaneous changes in junction temperatures of individual IGBTs.	As above.
4	Undervoltage lockout (Control supply failure)	This prevents any individual IGBT from linear region operation as soon as the gate supply voltage diminishes below a critical level.	As above.

overshoot—that is, the magnitude of the stray inductance and the rate of decay of current. High-power IGBT modules switch large current magnitudes over a short duration of time generating potentially destructive voltage transients. As these modules comprise several IGBT chips connected in parallel, each chip switches its separate share of load current at a rate *di/dt* governed by the gate drive circuit. The sum-total current and current fall rate *di/dt* in the external power circuit is obtained by adding together the individual currents and *di/dt* in each constituent IGBT chip. The situation becomes detrimental when a short-circuit current that is 5–10 times the normal rated current magnitude is switched off. The resultant *di/dt* is around several thousand amperes per microsecond.

For normal turn-on and turn-off switching, preventive measures adopted include (a) circuit layout improvement, for stray inductance minimization, and (b) choosing source capacitance with inherently low self-inductance. Properly selected decoupling capacitors (to avoid oscillations in the DC loop), connected across the module terminals, are a useful aid. Additionally, a voltage clamping circuit can be provided.

For protection against voltage transients produced during fault current turn-off, special protection circuits [15] are employed. These circuits are activated only upon sensing a fault current. Then the rate of fall of gate voltage is reduced, thereby decreasing the rate of decay of the fault current. This can be done by switching either (a) a high resistance in series with the IGBT gate or (b) a high capacitor in parallel with the input gate capacitance.

An important integration concept, "the rugged IGBT module," has emerged [16]. Needless to say that the primary design criteria for a power module are the minimization of conduction and switching losses and the maximization of ruggedness and reliability. Shortening of the DMOSFET channel length in the IGBT is a simple method to reduce the ON-state losses, but at the same time the IGBT becomes more prone to short-circuit failure. This vulnerability to short-circuit failure is removed by the integration of protection circuits in the IGBT module. The protection circuits are activated only when a malfunction takes place. Apart from the protection circuits, no other circuits are integrated. The user determines the switching and the forward operation of the IGBT through an external gate drive stage.

9.6 PARASITIC COMPONENTS OF IGBT MODULE PACKAGE

Figure 9.5 shows the parasitic components associated with an IGBT module package [17–19]. These consist of the inductances of the wires connecting the gate, emitter, and collector pads with the terminals on the package. These are denoted by L_G, L_E, and L_C, respectively. The resistance R_G is included to prevent gate oscillations.

Gate Side Effects. First, the gate inductance L_G is responsible for spurious turn-on or turn-off of the IGBT if the amplitude of the ringing exceeds the threshold voltage. Second, the construction of the IGBT module, by assem-

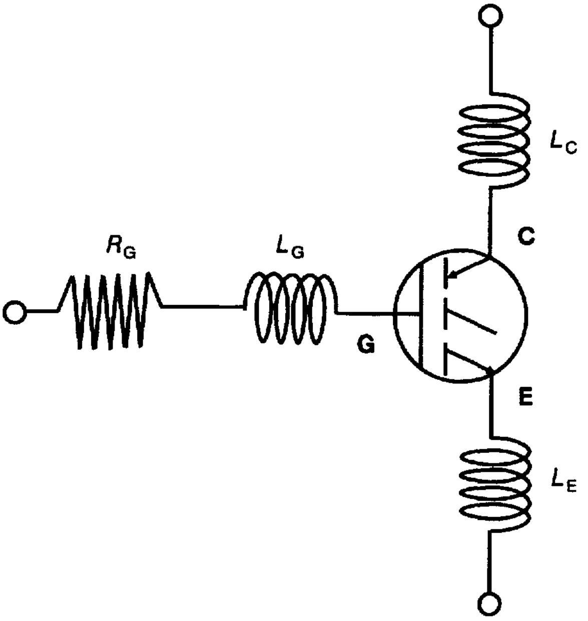


Figure 9.5 Circuit representation of IGBT with parasitic components.

bling a large number of chips, results in an extensive increase in the total chip area and, hence, the capacitance. The input capacitance is typically 80 nF at 10 V for 400 A IGBT, and the corresponding inductance is 50 nH. Together, these limit the bandwidth of the IGBTs response to gate drive circuits. Third, the gate resistance R_G, together with gate inductance L_G, accounts for introducing triggering delays between the different chips, comprising the module. This is due to the varying lengths of the wire bonds connecting the gate terminal to these chips. As the turn-on and turn-off of these chips varies, the magnitude of currents flowing in them, at a given instant, is different, subjecting them to different switching stresses, reducing the reliability of the module.

Collector Side Effects. The main collector side effect is the increase in output capacitance causing oscillations in the circuit; the collector inductance appears mainly as part of the turn-on snubber inductor. These oscillations are particularly troublesome during turn-on, when the equivalent output capacitance of the IGBT is very low (~ 200 pF at 600 V for a 400-A IGBT) because of high collector voltage at this time and the low quantity of charge in the drift region. High-frequency ringing during switching transients radiates electromagnetic interference (EMI). During turn-off the charge raises the capacitance to the level when the oscillations are adequately damped.

The emitter lead inductance L_E prolongs the turn-off of the IGBT. As the collector current begins to fall, a negative voltage acts across L_E by virtue of the *di/dt* effect, reducing the gate driving voltage applied to the IGBT (and hence its charge extraction capability), thereby lengthening the turn-off process. Collector and emitter lead inductances influence the switching pattern of the IGBT, hard- or soft-switching type, quite differently. During hard switching, considerable voltage stress and ringing take place. In zero-voltage switching (ZVS), L_E decreases the effective gate drive of IGBT and affects the turn-off process during the initial fall of collector current. Because the collector voltage is still very small, the effect on the losses is insignificant. di/dt is controlled to accomplish a softer turn-off, which is advantageous for the load. In zero-current switching (ZCS), the IGBT is still conducting as the current is reduced to zero. Although no significant power loss occurs due to the energy exchange between reactive components, oscillations arising from the resonance between the snubber capacitor and package inductance can cause severe electromagnetic interference. Thus, the IGBT performance deviates from the ideal case due to the above factors. Over and above, stray inductances produce large peak voltages during switching transients whereby the device operation may move out of the safe operating area.

9. 7 FLAT-PACKAGED IGBT MODULES

Flat- or press-pack encapsulation [20–24] offers several advantages over the traditional module package, as pointed out in Table 9.3. In a press-pack IGBT, the electrical contact is established by pressing the IGBT chips between two high-planarity conducting disks. An adequate stress relief layer is included to forebear the compression. The outermost poles are the *emitter* and *collector poles*, both made of nickel-plated copper. On the emitter side, a molybdenum washer and a nickel-plated copper foil are inserted between the emitter pole and the IGBT chip. Because the copper foil is a soft material, it ensures proper electrical contact and also enhances uniformity of pressure distribution. On the collector side, a molybdenum disk is placed between the collector pole and the IGBT chip. The IGBT chip, the foil, and the Mo washer assemblies are kept centered over the respective emitter pole probes with the help of a frame-centering device. This device also supports the gate contacts. For applying contact pressure to the pads, an external clamping system exerts force on the emitter and collector poles.

Figure 9.6 shows the 2.5-kV, 1-kA press pack IGBT module [24] for industrial and traction applications. This multichip module contains nine IGBT chips with three free-wheeling diodes, arranged in a matrix of 3 × 4, on a Mo base plate, with the IGBT chips connected to the outer gate bus through Si chip resistors. On the emitter side of the chips, the Mo metals are arranged to form the emitter contact. The chips are housed in a compact both-side-cooled hermetic, square, flat-packaged ceramic structure, having a size of $13.3 \times 11 \times 2$ cm^3, with pressure contacts. Low inductance is achieved inside the package because of the absence of Al wire bonding. For compactness, the unused area in the assembly is minimized. Edge termination consumes a great percentage of chip area. Any dead space in the package

Table 9.3 Comparison of Module Package with Flat Package

Serial No.	Module Package	Flat Package
1	The main disadvantage of this package is the high inductance of the wires connecting the chips with each other and the package terminals.	Offers the advantage of low inductance because of the absence of bond wires, with the connections made by "pressure contacts."
2	For good bonded joints, the wire should be small in diameter (typically 0.3 mm). The thin wire is also unable to conduct the heat away. So, the wire introduces significant electrical and thermal resistance.	Absence of wires overcomes this limitation.
3	Nonhermetic and one-side cooling structure.	Hermetic and both-side cooling structure.
4	The base plate is completely isolated making the cooling simple and lowering the cost.	Insulation of the IGBT and cooling are necessary, increasing the cost.
5	Poor power cycling capability.	Superior power cycling ability.
6	Undefined failure mode. Can be either open- or short-circuited.	Behaves as a short-circuit after destruction. This is helpful in applications where a redundant series connection of $(n+1)$ or $(n+2)$ devices are used because the converter continues to function, even if one or two IGBTs fail. These can be replaced during routine maintenance work.
7	Possibility of explosion during failure.	Explosions are avoided.
8	Less reliable.	More reliable.

also increases the useless area. Using a large chip area and thereby reducing the number of chips significantly lowers the area consumption by edge termination. Furthermore, replacing the circular package by a square package also decreases the dead space. The IGBT chip area is 2 cm^2. Divided into 36 independently working gate–emitter units, the current capability of a chip is 1.2 times the rated value. Thus the package can be repaired if up to four IGBTs fail. The saturation voltage of the IGBT module is 4.2 V at 1 kA, and the turn-off capability is more than $3 \times I_C$ under $V_{CC} = 1300$ V. Maximum thermal impedance of the IGBT portion is 15 K/kW. For double-side cooling, this thermal impedance is 0.4 times that of the module package. However, an oscillatory waveform is observed at the tail current period.

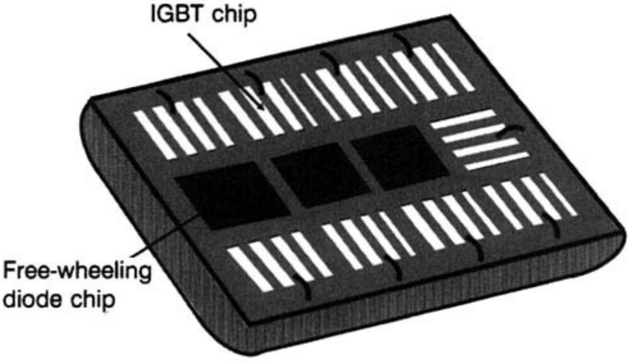


Figure 9.6 Element view of the (2.5 kV, 1 kA) press pack IGBT containing nine IGBT chips and three free-wheeling diodes.

These oscillations have been studied, and it has been found that they do not originate from the current capability differences between the multichips with the repair chips. Instead, due to the absence of the N-buffer layer in the NPT IGBTs used in the modules, the spreading of the depletion region deep into the N^- base produces a small difference in the capacitances of the multichips. Interaction of these capacitances with the small inductance between the multichips is the source of tail current oscillations. The oscillatory effects can be mitigated by increasing the inductance, but the advantage of a low parasitic inductance is lost. So, a high magnetic permeability material permalloy is included, surrounding the emitter contact metal, thereby preventing the interaction of currents amongst the IGBT chips. This ring has a large inductance at low currents but small inductance at high currents. Effectively, the inductance comes into play only for the duration of low currents. Thus the module has both low parasitic inductance and nonoscillatory characteristics.

9.8 DESIRABLE FEATURES AND RELIABILITY ISSUES OF IGBT MODULES

Basic requirements from a power module are as follows: (i) The electrical insulation of the IGBT chip from the heat sink is to be sturdy and secure. (ii) Thermal contact of the power dissipating IGBT chip to the heat sink requires a low resistance and high efficiency. (iii) The electrical lead wires interconnecting the IGBT chips among each other and with the external plugs require high conductivity. (iv) The encapsulation protecting the vulnerable IGBT chips and fixing the external leads must be strong.

Performance of an IGBT module is evaluated from both electrical and thermal viewpoints, and it must meet the anticipated electrothermal specifications [11]. Electrical parameters of interest are (i) the stray inductance in series with the chip, (ii) the capability of mounting a low-inductance bus to the terminals, and (iii) the capacitive crosstalk from one IGBT chip to another which causes an unwanted ON-state by charging the gate of an IGBT in OFF-state. The inductances set the minimum switching loss while the capacitance is responsible for radiation of electromagnetic interference.

Besides being economical, the module should have long life and high reliability. Price reduction is achieved by decreasing material and manufacturing costs and avoiding materials like molybdenum, which are difficult to machine. Alumina substrates are commonly used as a compromise between cost and optimal function. Longevity and reliability are assured by (i) keeping the temperature of the module low during operation and (ii) minimizing thermal cycling, which produces fatigue at the interfaces between different materials, due to difference of their coefficients of thermal expansion. Because these materials expand and contract at different rates, stresses are created at their interfaces by which the solder joints crack and the wires are dislodged from the bonds. At elevated temperatures, dendrite growth and impurity migration are hastened. Particular attention must be paid to the isolation of the base plate from the IGBT chip. This is enhanced by including both halves of the phase leg in one package. Moreover, it allows the mounting of modules, switching different phases, on a single heat sink. For safety, the heat sink is grounded.

Modules are used under different environmental conditions. Thermal stresses originate in a traction module—for example, from the frequent stoppage and running of transportation vehicles. Buses and trams have a stop-and-go period of around 1 minute; the period for short-distance passenger trains is several minutes, while that for fast passenger trains is 1 hour to several hours. Failure causes of IGBT modules have been the subject of intense study. Major determinants of the reliability of IGBT modules [25, 26], are described below:

(i) Effect of thermal cycling and mechanical shocks on the wire bonds causes many bonds to lift with associated increase in forward potential drop. This is referred to as bond-wire lift-off. A high-power IGBT module, containing approximately 450 wires with 900 wedge bonds, may fail due to disconnection of the wires, caused by heel cracks, bond liftoffs, and corrosion of wires. The hottest spots near the junction are subjected to maximum stress. The reliability parameter accounting for this mechanism is the junction temperature excursion ΔT_{j}. The wire bonds are particularly susceptible to fast thermal cycles, due to the small thermal mass of wires. Destruction of some wire bonds causes more current to flow through the smaller number of remaining bonds. A positive feedback process ensues with more current flowing through fewer bonds. Considerable improvement in reliability has been achieved by changes in the composition of wires, shape of the bonding tools, and bonding parameters, by metallization of the chips and leads, and by applying proper protective coatings on the wire bonding.

(ii) Thermal cycling causes bending of the whole assembly. This bending is the result of a "bimetal effect" due to the difference in coefficients of thermal expansion of the thick copper base and the DCB material. The results are (a) cracking of some interconnecting wire bonds and (b) poor thermal contact to the heat sink. Proper shaping reduces the bimetallic effect of the layered structure. A machined convex bow improves the heat transmission between the base plate and heat sink. Using a stiff material such as Al/SiC with low deviation of its thermal expansion coefficient to the ceramic will further alleviate the problem. Substitution of the Cu base plate by AlSiC plate having a smaller coefficient of thermal expansion enhances the thermal cycling capability by a factor of 5 from 3000 cycles.

(iii) Another important cause of failure is the solder between the Cu base plate and the substrate. The solder joining the base plate with the DCB substrate deteriorates upon thermal stresses, resulting in a rise in thermal resistance. Due to the difference in their coefficients of thermal expansion, thermal stress is created, which produces mechanical strain on the solder. The solder cracks upon repeated heavy load cycling, thereby increasing the thermal impedance between the chip and the base plate. Solder fatigue is mainly caused by slow thermal cycles. Due to the location of the largest solder joints between the case and the isolation of the modules, these joints are most likely to fail. The thermal parameter determining the solder fatigue is the case temperature excursion ΔT_c which depends on module construction and can be obtained from ΔT_j.

(iv) Migration of the grease, producing a nonuniform thermal resistance with consequent inhomogeneous temperature distribution.

(v) Replacement of the module package by press pack, employing only a few mechanical parts, results in enhanced reliability. In a press-pack multi-IGBT chip configuration [27], uniformity of power distribution among the parallel-connected IGBT chips, specification and conformance of dimensional and flatness tolerances, and proper stress/strain distribution are necessary for the various components such as the chip, foil, washer, and contacting probes. Temperature changes affect the contact pressure, because of differential deformation of the IGBT device, resulting from the mismatching of coefficients of thermal expansion, of different layers. For flat contacts, under elastic conditions, contact pressure gradients are noticed near the contact edges. As soon as the contact shear tractions overcome the frictional resistance between the bodies in contact, relative tangential displacements occur. Thermal cycling causes oscillatory surface sliding. During uniform heating and cooling cycles, contact pressure changes produce alternating asymmetric pressure distributions on the contact regions leading to fretting damage. Thus, the reliability of a press-packaged multi-chip IGBT depends on the application of satisfactory contact conditions during the assembly stage and their maintenance throughout the service life. For achieving this goal, the device sensitivity to damage initiation by thermomechanical fatigue and fretting, as caused by the different nonuniformities, must be eliminated.

(vi) Partial discharge (PD) and insulation resistance are important parameters for high-voltage IGBT modules. The discharge of metallized AlN ceramics in silicone gel equivalent to that used in IGBT modules has been investigated [28]. PD spectroscopy has shown that PD takes place at the interface between the ceramic and the silicone gel and at the rim of the copper metallization due to the high electric field. The shape of the metallization rim plays a crucial role with respect to the PD resistance. Poor etching or solder residues cause PD at low voltages around 5 kV. At higher voltages above 5 kV, discharge occurs at the interface between AlN substrate and the silicone gel along the path of high electric field strength emphasizing the need for silicone gel to adhere strongly to the ceramic substrate.

9.9 MODULE HEAT SINKS AND COOLING

Proper choice of the heat sink ensures a low junction temperature and hence a long life of the module. The three popular cooling methods are air cooling, liquid cooling, and evaporative cooling. For up to 100 kW of power dissipation, air cooling is preferred for economic reasons and simplicity. Above 100 kW, liquid or evaporative cooling is used for building compact systems and saving material costs.

Transference of heat across interfaces is expressed by the equation

$$\frac{dQ}{dt} = \alpha A \Delta T \tag{9.1}$$

where dQ/dt is the heat flow rate, α is the heat transfer coefficient, A is the area of the interface, and ΔT is the temperature difference across the interface. Maximum permissible junction temperature restricts the range of ΔT. Thus, the parameters α and A must be adjusted to increase the heat flow rate.

The α value for air-cooled systems is typically 5 W/m^2 K for free convection and 50 W/m^2 K for forced air flow. The air-cooled heat sinks usually have large fins attached to the base plate to increase the effective area. The size and shape of the heat sink must be chosen depending on the intended application. This is because the outer parts of a large heat sink, far away from the IGBT chip, may not contribute to cooling, reducing the sink efficiency.

The heat transfer coefficient for liquid coolants is an order of magnitude higher than air, enabling design compactness; but closed-loop systems, with pumps, are required. Additives such as ethyleneglycol or propyleneglycol extend the application range of water below 0°C. But water is corrosive and allows the use of only special Al alloys or stainless steels. Oils are inflammable, and fluorocarbons produce environmental hazards.

Regarding evaporative cooling, at high-power densities (10–100 W/cm^2), heat pipe cooling serves as an efficient heat spreading technique, distributing a high-power density on one side of the pipe, over a large surface area on the other side, where other methods can dissipate the heat. Here, a source of heat evaporates water or alcohol on one side of an enclosed space. Condensation of the vapors occurs on the other side of the pipe. By capillary forces, the condensed liquid returns to the heat source side and the cycle is repeated. Due to the high value of heat of vaporization, heat pipes serve as efficient heat removers. The disadvantages are: the mounting difficulty and a complicated cooling system. Another evaporative cooling method is immersion cooling in which the assembly is dipped directly in the liquid contained in a liquid bath. Heat produced in the device is carried away through the convection currents to the liquid surface. The vapors produced by evaporation of the liquid are condensed on a heat exchanger and fall back to the liquid bath where they are used again. Liquid nitrogen and fluorocarbons are the liquids widely used as they provide electrical isolation and stability.

9.10 MATERIAL REQUIREMENTS FOR HIGH-POWER IGBT MODULES

A high thermal conductivity is the prime necessity for power IGBT packages. Electrical insulation is another required property. But the same material may not fulfill both requirements. For instance, ceramic materials provide good

Table 9.4 Properties of Common Materials used in IGBT Modules

Serial No.	Material	Symbol/ Formula	Function	Thermal Conductivity (W/mK)	Young's Modulus (GPa)	Coefficient of Thermal Expansion (ppm/K)
1	Silicon	Si	Material for IGBT chip fabrication	150	130–190	2.6
2	Aluminum oxide (alumina)	Al_2O_3	Substrate material for module stack construction	30	370	5.5
3	Aluminum nitride	AlN	Substrate material for module stack construction	190	340	3.1
4	Beryllium oxide	BeO	Substrate material for module stack construction	270	345	6.3
5	Copper	Cu	Base plate in module stack	390	120	16
6	Aluminum	Al	Metallization	240	70	23
7	Gold–Si eutectic preform	97% Au, 3% Si	Used for die mounting	27	80	13
8	Lead–tin solder	63% Pb, 37% Sn	Used for die mounting	50	30	25
9	Thermal grease	—	Used for making good thermal contact between dry surfaces	0.4–1	—	—

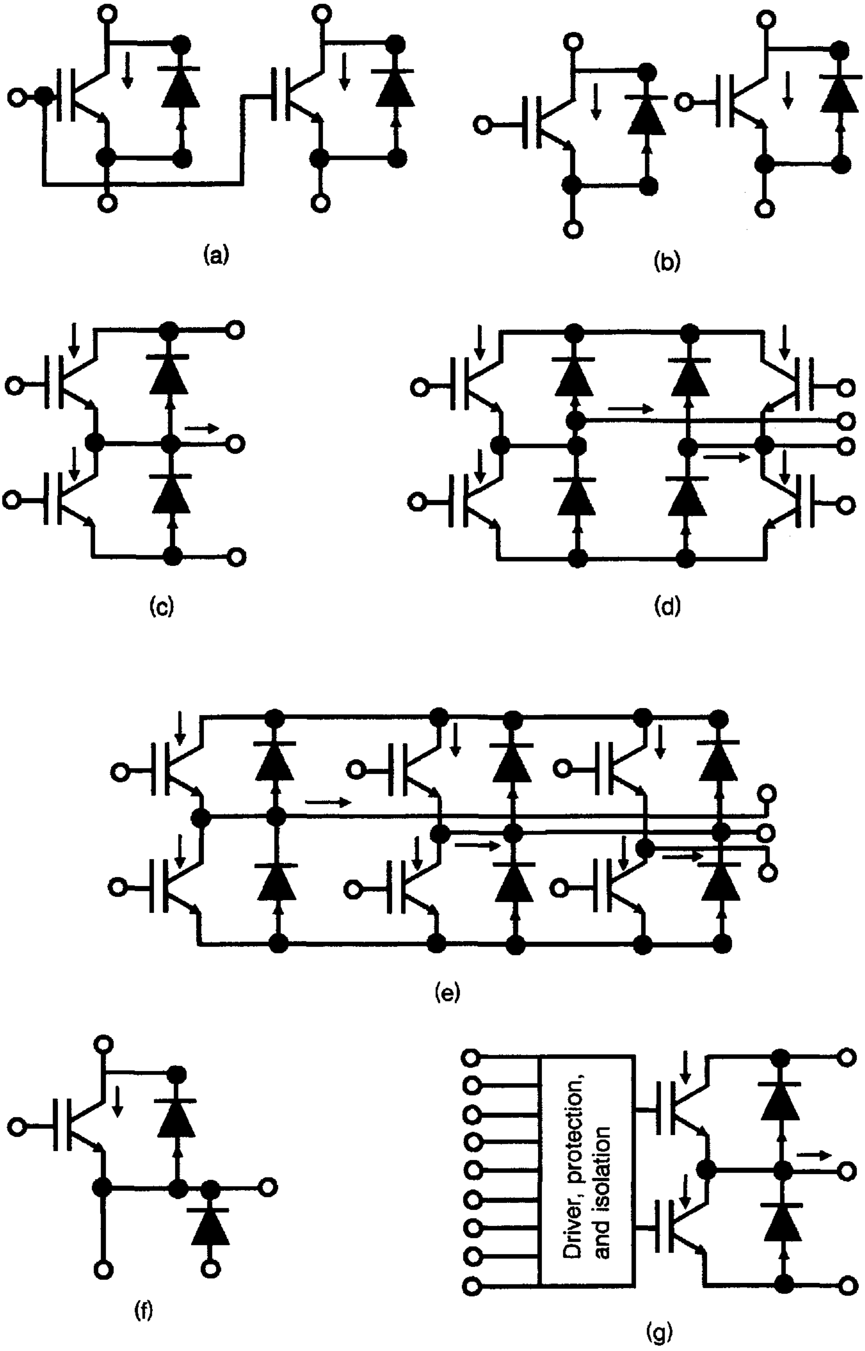
insulation with high dielectric strength but have a low thermal conductivity. Similar remarks apply to other materials used in IGBT packages. Additionally, the combination of different materials used in the package must be both thermally and mechanically compatible because thermomechanical stresses shorten the life span of the module. Young's modulus and coefficient of thermal expansion of these materials must, therefore, match closely. The effects of incompatible material combinations appear at the interfaces. High thermal resistances of soft solders or dry press contacts on the interface between the heat and the environment can increase the junction temperature beyond the permissible limit. The module design engineer takes all these factors into account using modern computational tools, such as finite element simulations, for studying the temperature distribution in the module.

Table 9.4 lists the important properties of materials used in an IGBT module.

9.11 STATE-OF-THE-ART AND TRENDS

The hybrid approach has gained immense popularity. Modules are constructed for the basic circuit building topologies, to simplify design for motor control, UPS, welding, electric vehicle, and other power switching applications. Figure 9.7 summarizes some types of commercial IGBT modules. A *bridge inverter* changes DC input to AC output without using a center-tapped load. *Half-bridge inverter* operates on half-cycle of the wave. Consisting of two bidirectional switches, it is a basic building block of voltage source DC-to-AC converters. Similarly, the *full-bridge inverter* operates over the complete cycle of the AC waveform, and the *three-phase inverter* operates over a three-phase AC input signal. *Chopper* is a circuit, which regulates the power flowing from a DC source to a DC load. It is used in regulated DC power supplies. Many more types of IGBT modules with various degrees of complexity are available, such as a module integrating three-phase inverter and three-phase rectifier in a single package for motor drive applications.

Monolithic integration of all functions is a futuristic goal [29]. Power modules with high levels of integration are designated as "smart power" or "intelligent power." Strictly speaking, they provide functions of self-protection and interfacing between power and control circuitry. It would be more apt to coin the term "integrated power modules" because an external microprocessor-based circuit essentially performs the intelligent control of a variable-speed drive. Versatility of a microprocessor defines the level where the intelligence of electronics begins. Present IGBT modules are divided into two categories: (a) the conventional module package derived from low-power technology and (b) the new press package based on the classical high-power diode or thyristor technology, Likewise, two types of power module suppliers have emerged, namely, firms having hybrid microelectronic facilities, wishing to expand into power electronics applications, and discrete device manufac-


Driver, protection, and isolation

(a) (b) (c) (d) (e) (f) (g)

Figure 9.7 Common IGBT modules; arrows indicate the directions of current flow paths. (a) Single module. (b) Dual module. (c) Half (H-) bridge inverter. (d) Full bridge inverter (or 4-pack module). (e) Three-phase bridge inverter (or 6-pack module). (f) Chopper module. (g) Smart IGBT module.

Table 9.5 Features of IGBT-Power Modules

Serial No.	Component of the Module	Features
1	Functionality	• Rectifier stage •• Inverter stage ••• Brake circuit •••• Temperature sensor ••••• Input logic, galvanic isolation between logic and power, gate drive, short-circuit, and undervoltage protection •••••• ESD protected digital interface
2	Fabrication technologies	Microelectronics, bare die handling, solder die attachment and wire bonding
3	Chip size and current density	Maximum size of the IGBT chip is 4.6 cm^2. IGBTs are specified at a current density of 80 A/cm^2 and free-wheeling diodes at a slightly higher current density of 100 A/cm^2.
4	Substrate	Direct bond copper (DBC) and insulated metal substrate (IMS) are used. The DBC substrate is more efficient in thermal dissipation. As thick Cu conductors are used, additional heat slugs are not necessary. IMS method uses thick heat-spreaders or metal slugs beneath the die due to the high thermal resistance of the dielectric.
5	Heat sink	This is attached to the substrate or baseplate forming the bottom of the module. Two lugs or holes in opposite short sides of the molding are used for bolting to the heat sink.
6	Encapsulant	Silicone gel is widely used. (silicone + hard epoxy) combination is also used. Silicone gel protects the circuitry from environmental contamination. It also provides voltage isolation. Plastic covers are screwed or glued to cover the surface. In the silicone and epoxy-filled module, the epoxy is used for holding the terminals tightly.
7	Top metallization	Aluminum (thickness ~ 4–6 μm). This is the minimum thickness recommended for heavy aluminum wire bonding, optimizing both bond reliability and cost.

Table 9.5 (*Continued*)

Serial No.	Component of the Module	Features
8	Wire bonding	Large diameter aluminum wire (~ 300 μm diameter) is used for interconnections. Ultrasonic bonding technique is employed. Multiple parallel wire bonds are used to increase the current carrying capability and to spread the current over the entire surface of the die for avoiding current crowding. As an example, the (3300 V, 1200 A) IGBT module contains 60 chips employing 450 wires with 900 wedge bonds.
9	Bottom metallization	Ti–Ni–Ag or Cr–Ni–Ag (thickness ~ 4 μm)
10	Die attach	SnAg3.5 solder (220°C), SnPb93 (280°C), SnPb37(180°C)
11	Input/output terminals	Ni-coated leads molded as part of the plastic frame. Exceptions may be mentioned like separate leadframes on plastic headers, soldered to the substrate and anchored to the package by the epoxy; and spring-loaded pressure connection, between direct-bonded copper and the user's printed circuit board.

turers using their own or externally acquired hybrid microelectronics capability to produce modules. The materials used in high-power modules should possess a high thermal conductivity, large insulation capability, excellent mechanical rigidity, and low (or matching) coefficient of thermal expansion.

IGBT modules commonly use NPT IGBTs because of simplicity in paralleling since they have a positive temperature coefficient of collector–emitter saturation voltage. Interesting features of an IGBT module are presented in Table 9.5. The specification sheet of an IGBT module is given in Table 9.6 for illustrative purposes.

Finally, it must be indicated that the higher manufacturing cost of advanced IGBT modules is compensated for by automated production techniques, resulting in a smaller number of components to be assembled, with reduced mounting, assembly, and wiring cost. Moreover, engineering costs are curtailed by the versatility achieved in terms of combinations of subsystems to fabricate systems and the decrease in the variety of components that must be assembled for system construction, due to software adaptation to particular requirements. Failure rates of advanced IGBT modules are minimized by the built-in self-protection schemes. Also, field bus connections are used for remote condition monitoring, so that maintenance and repair

Table 9.6 Typical IGBT-Module Specification Sheet

Serial No.	Parameter Symbol	Definition	Representative value
1	V_{CES}	Maximum collector–emitter voltage	1200 V
2	I_C	Maximum collector current (DC) at 25°C	200 A at 25°C
		Maximum collector current (DC) at 80°C	170 A at 80°C
3	$V_{CE(sat)}$	Collector–emitter saturation voltage	3.7 V
4	t_{ON}	Turn-on time	1.2 μsec
5	t_{OFF}	Turn-off time	5 μsec
6	E_{ON}	Turn-on energy per pulse at 125°C	30 mJ
7	E_{OFF}	Turn-off energy per pulse at 125°C	30 mJ
8	$P_{C(max)}$	Maximum collector power dissipation	1.1 kW
9	$R_{\theta JC}$	Junction-to-case thermal resistance	0.1 K/W

expenses are lowered. Thus a judicious consideration of all the advantages of system techniques, together with automated production techniques, decreases the overall cost of advanced drive systems.

REVIEW EXERCISES

9.1 Why must a large number of IGBT chips be paralleled for fabricating a high-current device? What restricts us from using a single large-area IGBT chip, housed in a hockey puck package, like a high-power diode or thyristor ?

9.2 Mention three additional functions, components, or circuits that are included in the module, besides a large number of parallel-connected IGBT chips for carrying very high currents? What is a smart power module?

9.3 Are the various IGBT chips in a module package connected by pressure contact or through wires? Is the module package a hermetic one?

9.4 Give two advantages of constructing power modules instead of relying on discrete components.

9.5 What are two types of substrates commonly used for building power modules? Name the varieties available in both types.

9.6 Why is a leadframe necessary in CTF technology? For what type of module is this method particularly useful?

9.7 How is the copper sheet bonded to the ceramic substrate for making DCB substrate? Give two reasons for the popularity of DCB substrates.

9.8 Draw the cross-sectional diagram of (a) DCB substrate and (b) polymer-insulated metal substrate. Explain the functions of the different layers in each scheme.

9.9 Explain how the IGBT chips are mounted on the module substrate, and how the chip-to-substrate and chip-to-chip interconnections are made?

9.10 Outline the basic concept of dielectric isolation. Why is it considered a complete isolation? Comment on the low heat transfer efficiency achieved by this method.

9.11 Briefly describe the creation of self-isolated regions. How is this kind of isolation obtained in MOSFET fabrication?

9.12 What are the main steps involved in producing PN-junction isolation? Cite one application.

9.13 What is the meaning of the term "BCD"? Highlight the features of the three generations of BCD.

9.14 Draw the circuit diagram of a three-phase bridge circuit in an IGBT module. What is the role of the free-wheeling diode, and what are its desirable features?

9.15 What are the four common fault conditions in a converter? What measures are adopted to protect a module from faulty conditions occurring during operation?

9.16 How is a switching voltage transient produced under normal operation of IGBT and under fault conditions? What steps are taken to protect the IGBT from normal and fault-induced voltage transients?

9.17 Explain the "Rugged Module" concept and point out how it allows achieving a trade-off between losses and reliability?

9.18 Provide a circuit diagram representation of the parasitic inductances associated with an IGBT in a module package. Explain how these parasitics degrade the performance of the IGBT?

9.19 Point out the advantages and pitfalls of the module package in comparison to the press pack. Why do tail current oscillations take place in the flat-package module? How are these oscillations reduced without affecting the parasitic inductance?

9.20 What performance characteristics are used for evaluation of an IGBT module? How does the ability to withstand thermal cycling determine the reliability of an IGBT power module?

9.21 You are provided with a DC supply of 120 V, a load resistance of 0.9 Ω, and two IGBTs of current ratings 80 A. Determine the resistance to be connected in series with each IGBT for ensuring a reasonable current sharing?

9.22 A 1500-V DC supply delivers power to a load. If the power conversion is performed by IGBTs and the available IGBTs have specifications of 300 V, 15 A, how will the steady-state voltage sharing be accomplished?

REFERENCES

1. V. Rumennik, Power Devices are in the Chips, *IEEE Spectrum*, Vol. 22, No. 7, July 1985, pp.42–48.
2. H. D. Lambilly and H. O. Keser, Failure Analysis of Power Modules: A Look at the Packaging and Reliability of Large IGBTs, *IEEE Trans. Components, Hybrids and Manuf. Technol*, Vol. 16, No. 4, June 1993, pp. 412–417.

3. K. Sommer, J. Göttert, G. Lefranc, and R. Spanke, Multichip High-Power IGBT Modules for Traction and Industrial Application, in *Proc. Conf. Rec. 7th European Conference on Power Electronics and Applications, EPE '97*, 8–10 September, 1997, Trondheim, Norway, EPE Association, Belgium, pp. 1112--116.
4. H. Brunner, M. Hierholzer, T. Laska, A. Post, and R. Spanke, 3300 V IGBT Module for Traction Application, in *Proc. Conf. Rec. 7th European Conference on Power Electronics and Applications, EPE '97*, 8–10 September, 1997, Trondheim, Norway, EPE Association, Belgium, pp. 1056–1059.
5. H. Brunner, M. Bruckmann, M. Hierholzer, T. Laska, and A. Porst, Improved 3.5 kV IGBT-Diode Chipset and 800 A Module Applications, in *Proc. Conf. Rec. Power Electronics Specialists Conference (PESC '96)*, 1996, IEEE, New York, pp. 1748–1753.
6. M. Hierholzer, R. Bayerer, A. Porst, and H. Brunner, Improved Characteristics of 3.3 kV IGBT Modules, in *Proc. Conf. Rec. 20th International Power Conversion and Intelligent Motion Conference (PCIM '97)*, 10–12 June 1997, Nürnberg, pp. 201–204.
7. G. Hilpert and T. Züllig, Integrated Power Module in IGBT Technology for Modular Power Traction Converters, in *Proc. Conf. Rec. 7th European Conference on Power Electronics and Applications, EPE '97*, 8–10 September 1997, Trondheim, Norway, EPE Association, Belgium, pp. 1106–1111.
8. Technical Information IGBT Module FZ1200R33KF1, EUPEC, 1998.
9. B. K. Bose, Energy, Environment and Advances in Power Electronics, *EEE Trans. Power Electron.*, Vol. 15, No. 4, July 2000, pp. 688–701.
10. S. Bernet, Recent Developments of High Power Converters for Industry and Traction Applications, *IEEE Trans. Power Electron.*, Vol. 15, No. 6, November 2000, pp. 1102–1117.
11. C. V. Godbold, A. Sankaran, and J. L. Hudgins, Thermal Analysis of High-Power Modules, *IEEE Trans. Power Electron.*, Vol. 12, No. 1, January 1997, pp. 3–11.
12. aavidthermalloy.com, Aavid Thermalloy, Concord, NH, USA, Products, Interface Materials.
13. B. Murari, F. Bertotti, and G. A. Vignola (Eds.), *Smart Power ICs: Technologies and Applications*, Springer-Verlag, Berlin, 1996, pp. 3–6.
14. G. Mazumdar and D. Medaule, A New Generation of Intelligent Power Devices for Motor Drive Applications, *5th International Conference on Power Electronics and Variable Speed Drives, PEVD'94*, 26–28 October, 1994, IEE, London, pp. 35–41.
15. R. Chokhawala and S. Shobhani, Switching Voltage Transient Protection Schemes for High Current IGBT Modules, *Ninth Annual Applied Power Electronics Conference and Exposition (APEC)*, Vol. 1, Febebruary 13–17, 1994, Orlando, FL, IEEE, New York, pp. 459–468.
16. S. Konrad and I. Zverev, Protection Concepts for Rugged IGBT Modules, *Euro. Power Electron. Drives (EPE) J*, Vol. 6, No. 3–4, December 1996, pp. 11–19.
17. K. Xing, F. Lee, and D. Boroyevich, Extraction of Parasitics within Wire-bond IGBT Modules, in *IEEE Applied Power Electronics Conference and Exposition (APEC)*, 1998, IEEE, New York, pp. 497–503.
18. P. R. Palmer, A. N. Githiari, and R. J. Leedham, Some Scaling Issues in the Active Voltage Control of IGBT Modules, *28th Annual IEEE Power Electronics Specialists Conference (PESC97)*, Record Vol. II, St. Louis, June 22–27, 1997, IEEE, New York, pp. 854–860.
19. M. Trivedi and K. Shenai, Parasitic Extraction Methodology for Insulated Gate Bipolar Transistors, *IEEE Trans. Power Electron.*, Vol. 15, No. 4, July 2000, pp. 799–804.
20. Y. Takahashi, T. Koga, H. Kirihata, and Y. Seki, 2.5 kV-100-A-stack IGBT, in *Proc 6th International Symposium on Power Semiconductor Devices and ICs (ISPSD'94)*, Davos, Switzerland, May/June 1994, pp. 25–30.

21. M. Hiyoshi, S. Yanagisawa, K. Nishitani, K. Kotaka, H. Matsuda, S. Teramae, and Y. Baba, A 1000 A, 2500 V Pressure Mount RC-IGBT, in *Proceedings 6th European Conference on Power Electronics and Applications, Seville, Spain, 19–21 September 1995 (EPE'95)*, EPE Association, Belgium, pp. 1.051–1.055.
22. H. Y. Takahashi, K. Yoshikawa, M. Soutome, T. Fujii, M. Ichijyou, and Y. Seki, 2.5 kV-1000A Power Pack IGBT (High Power Flat-Packaged RC-IGBT), in *Proceedings International Symposium on Power Semiconductor Devices and ICs (ISPSD'96)*, 1996, Hawaii, pp. 299–302.
23. Y. Takahashi, T. Koga, H. Kirihata, and Y. Seki, 2.5 kV-100A Flat-Packaged IGBT (Micro-stack IGBT), *IEEE Trans. Electron Devices*, Vol. 43, No. 12, December 1996, pp. 2276–2282.
24. Y. Takahasi, K. Yoshikawa, M. Soutome, T. Fujii, H. Kirihata and Y. Seki, 2.5 kV–1000 A Power Pack IGBT (High Power Flat-Packaged NPT Type RC-IGBT), *IEEE Trans. Electron Devices*, Vol. 46, No. 1, January 1999, pp. 245–250.
25. A. Hamidi, G. Coquery, and R. Lallemand, Reliability of High Power IGBT Modules Testing on Thermal Fatgue Effects due to Traction Cycles, in *Proc. Conf. Rec. 7th European Conference on Power Electronics and Applications, EPE '97*, 8–10 September, 1997, Trondheim, Norway, EPE Association, Belgium, pp. 3118–3122.
26. Th. Schütze, H. Berg, and M. Hierholzer, Further Improvements of Reliability of IGBT Modules, in *Proc. Conf. Rec. of the IEEE Industry Applications Society (IAS) Annual Meeting*, IEEE, New York, 1998, pp. 1022–1025.
27. A. Pirondi, G. Nicoletto, P. Cova, M. Pasqualetti, M. Portesine and P. E. Zani, Thermo-mechanical Simulation of a Multichip Press-Packed IGBT, *Solid State Electron.*, Vol. 42, No. 12, December 1998, pp. 2303–2307.
28. G. Mitic and G. Lefranc, Localization of Electrical Insulation and Partial Discharge Failures of IGBT Modules, *IEEE Trans. Indust. Appl.*, Vol. ED-38, January/February, 2002, pp. 175–180.
29. J. Flannery, M. Meinhardt, P. Cheasty, P. Byrne, S. C. Ó. Mathúna, O. Slattery, K. Briggs,
P. Barrass, and B. Drury, State-of-the-Art of Integrated Power Modules (IPMs) for 0.75 kW and 2 kW Drive Applications, *IEEE Applied Power Electronics Conference and Exposition (APEC)*, 1999, IEEE, New York, pp. 657–663.

10

NOVEL IGBT DESIGN CONCEPTS, STRUCTURAL INNOVATIONS, AND EMERGING TECHNOLOGIES

Since the introduction of IGBTs, efforts have been made to improve the device characteristics by optimizing the conduction and switching losses to achieve the best trade-off. As a result, device characteristics have reached the theoretical limit. Further improvements are possible by structural modifications such as incorporation of an additional gate, inclusion of extra P layer, and so on. In this chapter, some novel IGBT structures are reviewed. We will also deal with the improvements in existing IGBT structures, focusing on high-power and high-frequency applications. Based on a physical understanding of existing structures and their shortcomings, various researchers have carried out extensive multidimensional computer simulations, enabling them to investigate the conduction, blocking, and switching characteristics of the proposed structures and to comprehensively study their electrical performance. Many of these structures have been experimentally demonstrated in the laboratory, some are in their early developmental stage, and some are awaiting prototyping.

10.1 TRADE-OFF BETWEEN ON-STATE VOLTAGE DROP AND SWITCHING LOSSES

The general notion is that the greater the charge stored in the base during ON-state, the smaller the forward drop and the larger the turn-off energy

---

The Insulated Gate Bipolar Transistor (IGBT) Theory and Design, By Vinod Kumar Khanna.
ISBN 0-471-23845-7 © 2003 Institute of Electrical and Electronics Engineers

loss. This is not necessarily true, because these effects depend upon the base carrier distribution profile [1(a)]. Increasing the carrier distribution near the emitter is advantageous, and an optimum carrier distribution profile can be found to achieve the best trade-off. Under the simplifying assumptions of a long carrier lifetime in the N^- base, a one-dimensional carrier distribution in the base, a strong gate drive voltage, and clamped inductive load switching conditions, the following parameters are calculated [1(a)].

(i) The electron current at the collector junction [Appendix 10.1, Eq. (A10.1.4)]:

$$I_{\mathrm{nW_b}} = \frac{b}{1+b} I_{\mathrm{t}} + \frac{qAD_{\mathrm{a}}(p_{\mathrm{w}} - p_0)}{W_{\mathrm{b}}} \tag{10.1}$$

where I_{t} is the total device current, b is the electron-to-hole mobility ratio, A is the device area, D_{a} is the ambipolar diffusivity, p_{w} is the emitter side carrier concentration, p_0 is the collector side carrier concentration, and W_{b} is the base width.

(ii) The instantaneous base charge at time t [Appendix 10.2, Eq. (A10.2.1)]:

$$Q_{\mathrm{base}}(t) = Q_0 - Q' \tag{10.2}$$

where Q_0 is the initial stored charge in the base and Q' is the transient base stored charge.

(iii) The depletion layer width $W_{\mathrm{dep}}(t)$ at time t from the transient base charge [Appendix 10.2, Eqs. (A10.2.1) and (A10.2.3)]:

$$Q_{\mathrm{base}}(t) = Q_0 - \frac{qA}{2}\{(p_{\mathrm{t0}} + p_{\mathrm{t}})W_{\mathrm{dep}}(t) + (p_{\mathrm{t0}} - p_{\mathrm{t}})L_{\mathrm{C}}\} \tag{10.3}$$

where [Appendix 10.2, Eqs. (A10.2.4) and (A10.2.5)]

$$p_{\mathrm{t}} = p_{\mathrm{w}} + \frac{(p_0 - p_{\mathrm{w}})\{W_{\mathrm{dep}}(t) + L_{\mathrm{c}}\}}{W_{\mathrm{b}}} \tag{10.4}$$

$$p_{\mathrm{t0}} = p_{\mathrm{t}}|_{W_{\mathrm{dep}}(t)=0} \tag{10.5}$$

and [Appendix 10.2, Eq. (A10.2.2)]

$$Q_0 = \frac{qA}{2}\{W_{\mathrm{b}}(p_0 + p_w) - p_w L_{\mathrm{c}}\} \tag{10.6}$$

L_{c} is a constant distance in transient carrier distribution.

(iv) The collector voltage V_C at time t from the depletion layer width at time t [Appendix 10.3, Eq. (A10.3.4)]:

$$W_{\text{dep}}(t) = \sqrt{\frac{2\varepsilon_0 \varepsilon_s V_C(t)}{q(N_D + p_m)}} \tag{10.7}$$

where ε_s is the relative permittivity of Si, $V_C(t)$ is the collector voltage at time t, N_D is the N^--base concentration, and p_m is the mobile hole concentration.

(v) Modulated base resistance [Appendix 10.4, Eq.(10.4. 5)]:

$$R_{\text{base}} = \frac{W_b}{qA(\mu_n + \mu_p)} \frac{\ln p_0 - \ln p_w}{p_0 - p_w} \tag{10.8}$$

where μ_n and μ_p are the electron and hole mobilities.

(vi) ON-state voltage drop [Appendix 10.5, Eq. (A10.5.10)]:

$$V_{\text{ON}} = R_{\text{base}} I_t + V_{\text{MOS}} + \frac{kT}{q} \ln\left(\frac{p_0^2}{n_i^2}\right) \tag{10.9}$$

Now the energy losses due to the rising collector voltage are given by [Appendix 10.6, Eq. (A10.6.1)]

$$E_{\text{vr}} = I_t \int_0^{t_{\text{vr}}} V_C(t)\, dt \tag{10.10}$$

where t_{vr} is the time when the collector voltage $V_C(t)$ reaches the switching rail voltage. Including the carrier recombination in the base, the energy loss due to the current tail is [Appendix 10.6, Eq. (A10.6.2)]

$$E_{\text{tail}} \cong V_{\text{DC}} \cdot Q_{\text{base}}(t)\big|_{V_C(t)=V_{\text{DC}}} \tag{10.11}$$

Applying Eqs. (10.8)–(10.11), the trade-off between the ON-state voltage and turn-off loss incurred by the IGBT is calculated for specified p_w, p_0, V_{DC}, and I_t. This analysis shows that the optimum carrier concentration is one that has higher carrier concentration on the emitter side than on the collector side. The problem then reduces to finding the IGBT structure, which will give the optimum carrier distribution. The trench IGBT structure provides such carrier distribution. It can achieve this because the carrier injection from the N^+ accumulation region into the N^- base in the trench IGBT structure is much larger than that in the planar IGBT. Further increase in the emitter-side carrier concentration is accomplished by increasing the trench gate width and therefore accumulation gate area. For the trench IGBT, the controlling parameter for the injection efficiency from the emitter side is the trench gate width but that from the collector side is the P^+ implant dose. To illustrate, consider two punchthrough IGBT structures X and Y having trench gate widths of 5 μm and 10 μm with P^+ collector doping concentrations of 8×10^{17} and 3×10^{17} cm^{-3}, respectively [1(a)]. IGBT B has a higher emitter injection efficiency but lower collector injection efficiency. IGBT B will therefore provide a better trade-off between conduction and switching losses. But it must be pointed out that a broader trench gate is problematic for polysilicon refilling.

An NPT-IGBT structure giving weak collector injection efficiency is the *transparent collector structure*. Here, a precisely controlled acceptor impurity is implanted into the backside of the IGBT, yielding a thin and low-doped collector. In the case of the PT-IGBT, a highly doped N^+ buffer layer is employed to decrease the collector injection efficiency. Localized lifetime killing near the collector also increases the recombination in the vicinity of collector junction, thereby bringing down the collector injection efficiency. Use of multiple methods of controlling both the collector and emitter side injection efficiencies is helpful in designing optimum structures.

10.2 PARALLEL AND COUPLED PIN DIODE–PNP TRANSISTOR MODEL OF CARRIER DISTRIBUTION IN THE ON-STATE OF TRENCH IGBT

During the forward conduction state, the carrier distribution in a trench IGBT is the resultant effect of the carrier profiles in the PIN diode (formed by P^+ collector/N^- base/P base) and the PNP transistor (between P^+ collector/N^- base/N^+ accumulation layer) [1(b)]. The relative proportion of contributions from the PIN diode and PNP transistor is governed by the ratio of the total accumulation layer length l (horizontal length l_{horiz} + vertical length l_{vert}) to the cell width d, called the design aspect ratio l/d (Fig. 10.1). The overall effect of the PIN diode–PNP transistor combination in trench IGBT is treated in terms of two effects, namely the *parallel effect*, a geometrical phenomenon, arising from the parallel action of two components, and the *coupled effect*, describing the dependence of carrier distribution in the PIN diode region of the N^- base, in the IGBT, on the static current gain of PNP transistor α_{PNP} [1(b)].

When the design aspect ratio is increased, the excess carrier concentration is enhanced in both dimensions, reducing the forward potential drop. However, the switching characteristics are not affected because immediately upon cessation of the channel and accumulation layer, the electron injection stops. The carrier concentration falls down on the emitter side. The rapidly moving space-charge region on the emitter side sweeps away the excess holes towards the emitter short contact. But the recombination process, deeper inside the N^- base, is slower. Thus the PIN diode effect on the emitter side does not influence the turn-off behavior of the trench IGBT.

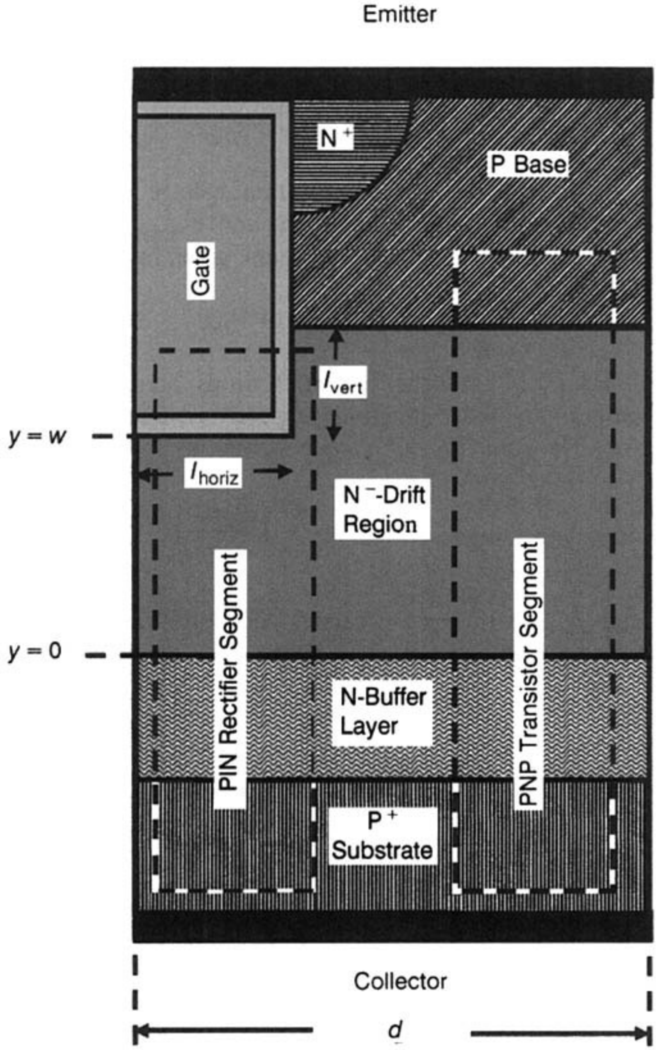

Emitter
N+
P Base
Gate
l_{vert}
y = w
l_{horiz}
N⁻-Drift Region
PIN Rectifier Segment
PNP Transistor Segment
y = 0
N-Buffer Layer
P+ Substrate
Collector
d

Figure 10.1 Trench-gate IGBT structure under analysis.

Defining an optimal trench IGBT as one with a large cell density having $l \gg d$, the parallel effect reduces to the action of PIN diode alone, because the PIN diode area is much larger than PNP transistor area. But the peak concentration of the PIN diode on the emitter side depends on the PNP transistor. The hole current traveling vertically upwards is decomposed into two components at the start of the trench gate. One component recombines in the accumulation layer, while the other flows through the P base to the emitter short contact. The PNP transistor causes a lower concentration p_w at the accumulation-layer/N$^-$-base interface, which is accounted for by the coupled effect. Based on this analysis, the saturation voltage of IGBT is expressed as

$$V_{\mathrm{CEsat}} = V_{\mathrm{DSsat}} + V_{\mathrm{PINsat}} \tag{10.12}$$

where V_{DSsat} is the drain–source saturation voltage of the MOSFET expressed by the usual MOSFET equation, and V_{PINsat} is the voltage drop across the PIN diode in the saturation region, given by

$$V_{\mathrm{PIN}} = V_{\mathrm{JC}} + V_{\mathrm{N^-}} + V_{\mathrm{JAN^-}} \tag{10.13}$$

where V_{JC} and $V_{\mathrm{JAN^-}}$ are the potential drops across the collector and accumulation-layer/N$^-$-base junction, respectively, and $V_{\mathrm{N^-}}$ is the N$^-$-base potential drop. The sum $(V_{\mathrm{JC}} + V_{\mathrm{JAN^-}})$ is written as

$$V_{\mathrm{JC}} + V_{\mathrm{JAN^-}} \cong \frac{kT}{q}\left(\frac{p_0 p_{\mathrm{w}}}{n_{\mathrm{i}}^2}\right) \tag{10.14}$$

where p_0 and p_{w} are the excess carrier concentrations in the N$^-$ base at the collector and emitter ends. The expression for p_{w} is [Appendix 10.7, Eq. (A10.7. 29)]

$$p_{\mathrm{w}} = \frac{D}{2Lh_{\mathrm{w}}}\left[-\coth(w/L) + \sqrt{\coth^2(w/L) + \frac{4h_{\mathrm{w}}\tau}{qD}\left(\frac{1}{1+b} - \alpha_{\mathrm{PNP}}\right)J + \frac{4h_{\mathrm{w}}Lp_0}{D\sinh(w/L)}}\,\right] \tag{10.15}$$

where D is the ambipolar diffusivity, w is the distance marked in Fig. 10.1, L is the ambipolar diffusion length, h_{w} is the recombination constant of the accumulation layer junction, w is the N$^-$-base thickness, b is the ratio of electron and hole mobilities, α_{PNP} is the static current gain of the PNP transistor, J is the ON-state collector–emitter current density, and p_0 is the excess carrier concentration in the N$^-$ base on the collector side. The potential drop $V_{\mathrm{N^-}}$ is given by [Appendix 10.8, Eqs. (A10.8.10) and (A10.8.11)]

$$V_{\mathrm{N^-}} = \frac{J\sin h(w/L)}{q(\mu_{\mathrm{n}} + \mu_{\mathrm{p}})} \times \int_0^w \frac{dy}{p_0\sinh\left(\dfrac{w-y}{L}\right) + p_{\mathrm{w}}\sinh(y/L)} + \left(\frac{kT}{q}\right)\frac{D_{\mathrm{n}} - D_{\mathrm{p}}}{D_{\mathrm{n}} + D_{\mathrm{p}}}\ln\left(\frac{p_0}{p_{\mathrm{w}}}\right) \tag{10.16}$$

where μ_n and μ_p are the electron and hole mobilities, and D_n and D_p are the corresponding diffusivities.

The above model shows that a high l/d ratio causes large conductivity modulation on the emitter side by virtue of the PIN rectifier effect, without sacrificing the switching speed. Because the PIN diode is a two-sided injection device, the above implies that the trench IGBT is a double-sided injection fast-switching device.

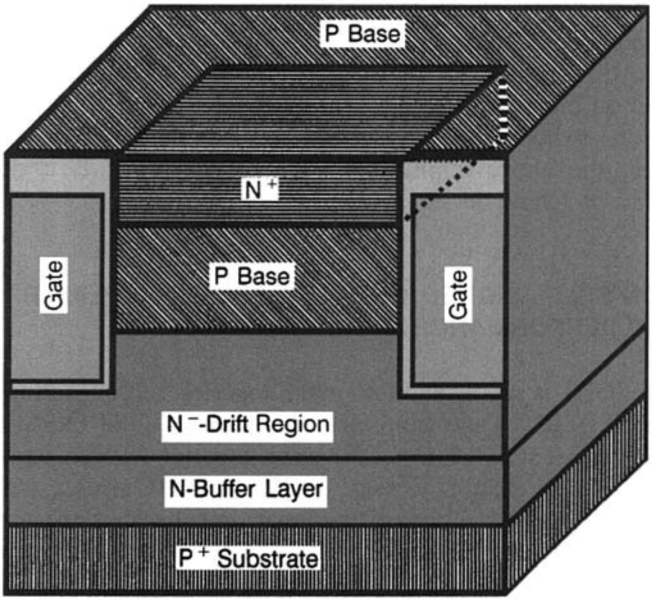
10.3 NON-SELF-ALIGNED TRENCH IGBT FOR SUPERIOR ON-STATE PERFORMANCE

The trench IGBT is subdivided into two categories: self-aligned trench IGBT (SA-IGBT) (Fig. 10.2b) and Non-self-aligned trench IGBT (NSA-IGBT) (Fig. 10.2a) [2]. The SA-IGBT is expected to be superior to NSA-type because it has a higher MOS channel density. But NSA-IGBT gives a better performance in terms of lower forward potential drop because the contact resistance of the N^+ emitter is smaller in this variety. This difference originates from the fact that the contact area to the N^+ emitter is larger for the NSA-IGBT. In an SA-IGBT, the dopant concentration in the semiconductor surface decreases laterally, in the electrical contact region joining the contact metal with the semiconductor surface, leading to a higher contact resistance. Additionally, the encroachment of a bird's beak (Fig. 10.2c) decreases the contact area to the N^+ emitter and allows the contact to be made to the emitter only in a region of high resistivity. Therefore, fabrication of high-performance SA-IGBTs calls for removal of the horizontal bird's beak.

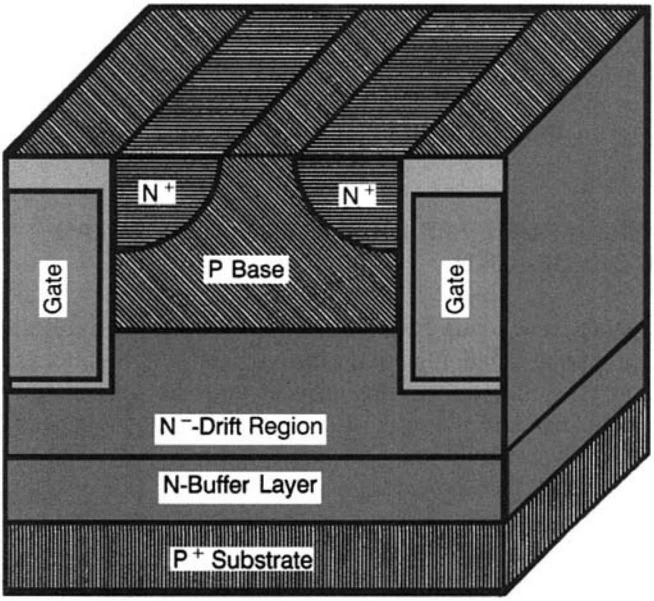
10.4 DYNAMIC N-BUFFER INSULATED GATE BIPOLAR TRANSISTOR (DB-IGBT)

This is a double trench-gate IGBT that can self-adjust the collector injection efficiency, providing a high injection efficiency in the ON state, and providing low injection efficiency during the turn-off and short-circuit conditions [3]. Thus a smaller ON-state potential drop is obtained together with high-speed switching. Cross-sectionally, the DB-IGBT resembles the single trench-gate NPT-IGBT, except that an additional trench gate G_2 is inserted on the collector side, as shown in Fig. 10.3a.

The width w and the doping concentration N_D of the mesa region between the two sides of the trench gate G_2 determine the electrical behavior of the DB-IGBT. The working of the gate G_2 is understood by applying a negative voltage to this gate, higher than the threshold voltage. Figures 10.3b and 10.3c show the energy band diagrams for two structures—for example, a two-side gate with large spacing w and a two-side gate with small spacing w. For the two-side gate with large width $w > 2X$, the

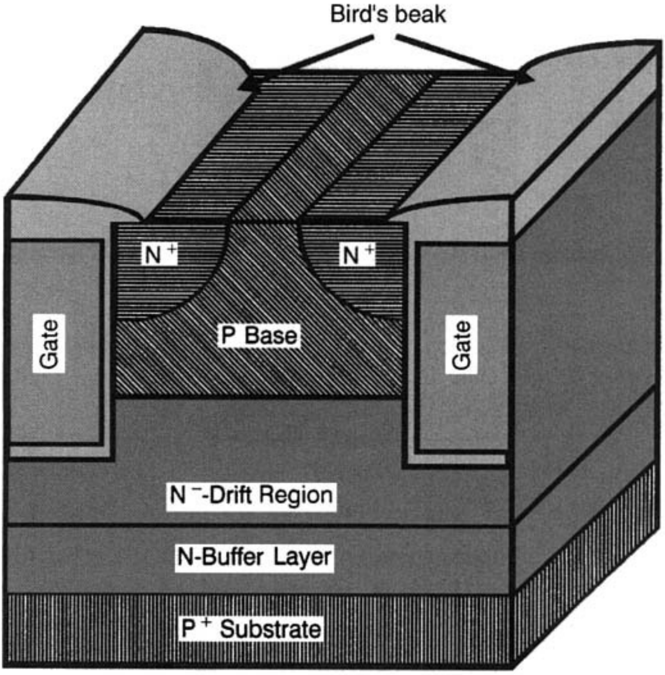

P Base
N+
P Base
Gate
Gate
N−-Drift Region
N-Buffer Layer
P+ Substrate

(a)


N+
N+
P Base
Gate
Gate
N−-Drift Region
N-Buffer Layer
P+ Substrate

(b)

Figure 10.2 Cross section of trench IGBTs. (a) Non-self-aligned trench IGBT. (b) Self-aligned trench IGBT. (c) Self-aligned trench IGBT showing the bird's beak effect.



(c)

Figure 10.2 (*Continued*).

depletion region extends in silicon up to the maximum depletion width X determined by the relative permittivity of Si, its doping concentration, and its Fermi potential. Because the depletion regions from the opposite sides do not merge, the remaining silicon portion between the depletion regions from the two sides is shown. But for the two-side gate with small width $w < 2X$, the mesa region is fully depleted.

The bias on gate G_2, with respect to the collector, controls the operation of DB-IGBT. Because the polysilicon gate G_2 is heavily doped, its Fermi level is coincident with the bottom of the conduction band. At the same time, the Fermi level of the lightly doped N^- base is located far away from the conduction band. Also, its work function is higher than that of G_2, and there exists a built-in potential difference across the P-collector/N^--base junction. Hence, an electron accumulation layer is formed at the interface near gate G_2 in the N^- base even when $V_{G2} = 0$ V. Effectively, a weak buffer zone is created in the mesa region. Upon biasing G_2 positively, the electron accumulation increases enhancing the N-buffer zone. A high-voltage withstanding capability is provided by the electric field shielding of the trench, as well as the electron accumulation in the mesa region, even when $V_{G2} = 0$ V. Thus the

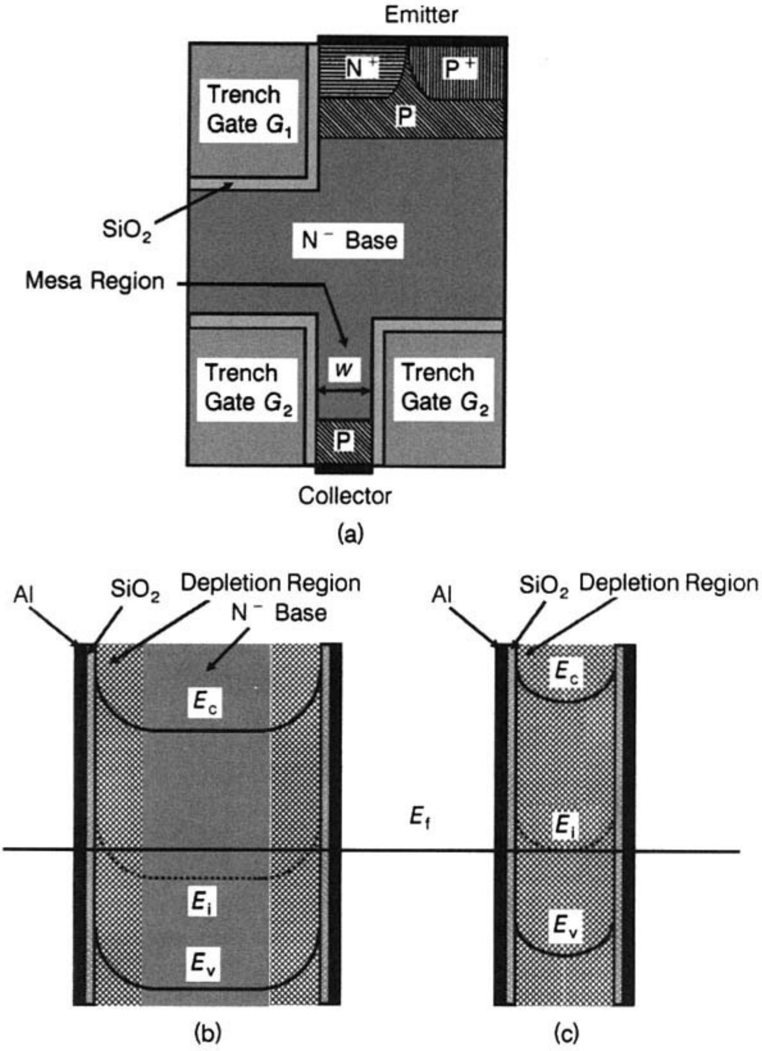

Emitter
N+
P+
P
Trench Gate G1
SiO2
N- Base
Mesa Region
Trench Gate G2
w
Trench Gate G2
P
Collector
(a)
Al
SiO2
Depletion Region
N- Base
Ec
Ei
Ev
Ef
Al
SiO2
Depletion Region
Ec
Ei
Ev
(b)
(c)

Figure 10.3 DB-IGBT cross section and associated energy band diagrams. (a) Half-cell of DB-IGBT. (b) Energy-band diagram with a wide spacing *w*. (c) Energy-band diagram with a narrow spacing *w*.

blocking characteristics are obtained. In the ON state, the gate G_2 is biased at a negative potential, -15 V. Then an inversion layer (hole density) is formed in the mesa region and at the bottom of the trench in the N^- base of IGBT. This inversion layer serves as a thin and heavily doped P^+ region, which injects holes into the N^--drift region like the P^+ collector and therefore supplements the injection efficiency of the collector. Thus, the

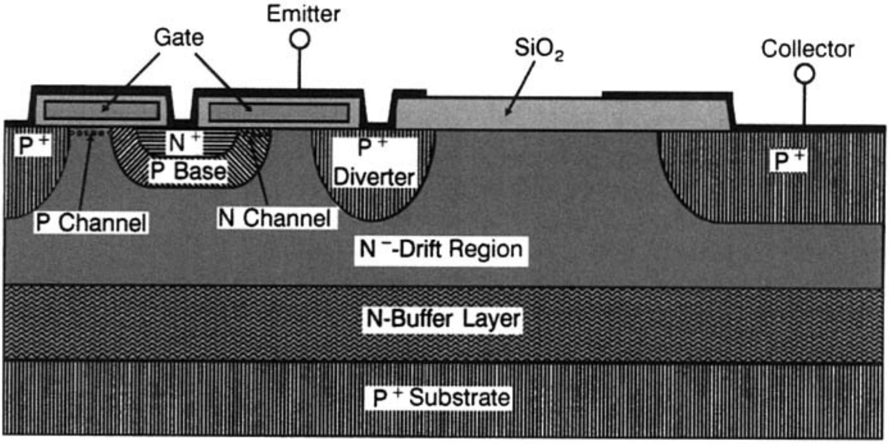
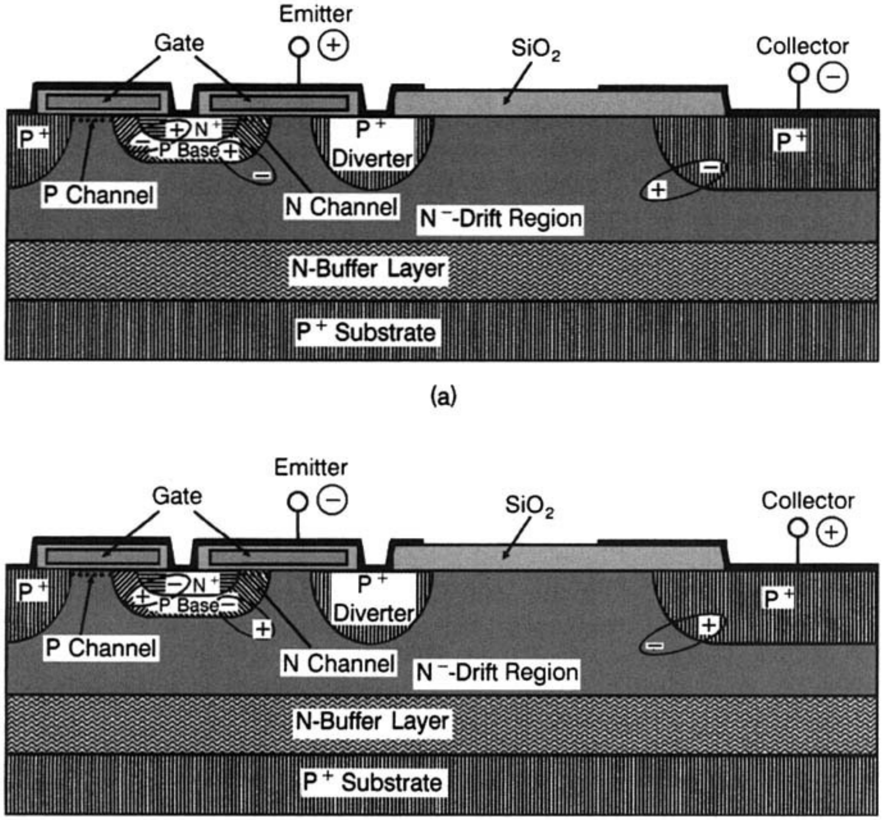


Figure 10.4 Lateral IGBT (with P$^+$ diverter region) having reverse blocking capability for AC power control.

injection efficiency of the P$^+$ collector is considerably enhanced by the inversion layer and can be controlled by monitoring the negative potential applied to the gate G_2. It may be recalled that in identical fashion, the application of a negative voltage to gate G_2 had provided the blocking action.

10.5 LATERAL IGBT WITH REVERSE BLOCKING CAPABILITY

To improve the trade-off between forward voltage drop and forward blocking voltage, the reverse blocking capability of IGBT has to be sacrificed. In applications where reverse blocking capability is required, a diode is connected in series with the IGBT to support the reverse voltage. But this practice increases the forward drop and the associated power losses. A lateral IGBT structure with reverse blocking capability is shown in Fig. 10.4. This IGBT structure [4] is useful for integration in a power IC for AC applications. Here, two P$^+$-diverter regions are included, one on each side of the double-diffused junction of P base and N$^+$ emitter. A single gate electrode controls both the N-channel MOSFET, formed by the DMOS structure, and the P-channel MOSFET due to the P$^+$ region, as shown in the diagram. Without application of any voltage to the gate, the device blocks current flow in both directions, as can be readily understood from the biasing of the different junctions in the device (Fig. 10.5). When a positive voltage is applied to the gate, the N-channel MOSFET is turned on. Then electron injection takes place from the N$^+$ emitter into the N$^-$-drift region. At the same time, hole injection occurs from the P$^+$ region at the collector. These holes reach the P$^+$ diverter and then through the N$^-$-drift region into the P


Emitter
Gate
SiO$_2$
Collector
P$^+$
N$^+$
P Base
P$^+$
Diverter
P Channel
N Channel
N$^-$-Drift Region
N-Buffer Layer
P$^+$ Substrate
Emitter
Gate
SiO$_2$
Collector
P$^+$
N$^+$
P Base
P$^+$
Diverter
P Channel
N Channel
N$^-$-Drift Region
N-Buffer Layer
P$^+$ Substrate

(a)

(b)

Figure 10.5 Lateral IGBT (with P$^+$ diverter region): (a) Positive potential on emitter and negative potential on collector of IGBT and (b) negative potential on emitter and positive potential on collector of IGBT.

base. At low current densities, the device works as an IGBT. But at high current densities, when the voltage drop across the P-base/N$^+$ emitter junction exceeds the built-in potential, the PNPN thyristor is turned on. Due to the regenerative feedback mechanism, heavy injection of carriers occurs, minimizing the forward voltage drop across the device. This is the thyristor mode of operation. To turn off the device, a negative bias is applied to the gate electrode when the P-channel MOSFET is activated. Then a shunt path for the holes is created connecting the P-base region and the P$^+$-inverter region. Thus the thyristor action is quenched and the device is switched off. The forward conduction drop of this device was 6.5 V at 100 A/cm^2, and the blocking voltages were 600 V in both the directions.

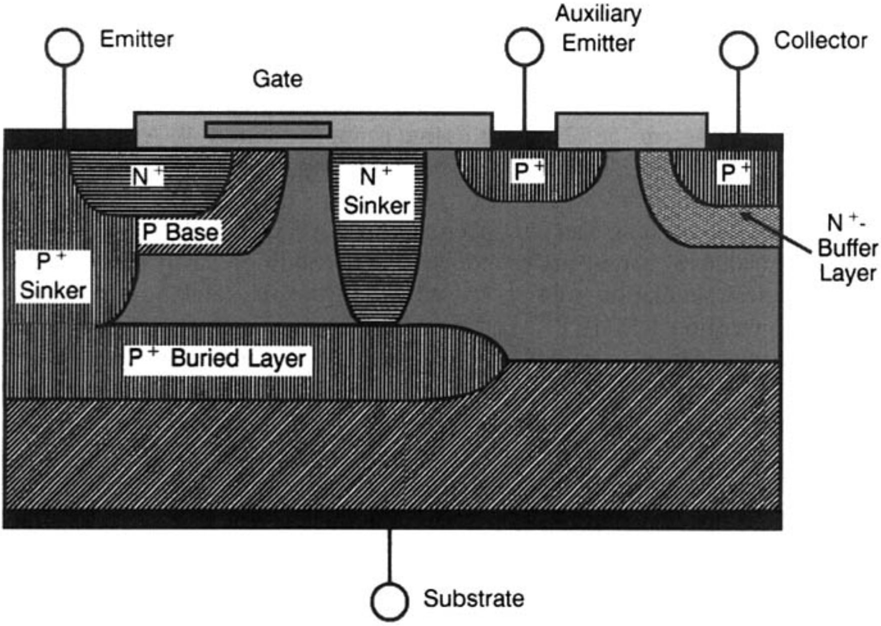


Figure 10.6 Lateral IGBT for high-temperature operation.

10.6 LATERAL IGBT WITH HIGH-TEMPERATURE LATCH-UP IMMUNITY

Figure 10.6 presents the cross-sectional view of this IGBT structure [5]. Attention is drawn to the inclusion of the auxiliary emitter, N^+ sinker and P^+ buried layer in the structure. During the nonconducting state, the region between the active channel and the auxiliary emitter is at low potential so that the highly doped sinker region does not degrade the device blocking voltage. But the sinker region enhances the forward operation of the device in two ways. First, it produces an electric field due to the carrier gradient avoiding the hole flow through the P base. As a result, the holes are diverted toward the buried layer and the auxiliary emitter. Thus, the hole current flow through the P base, causing latchup, is reduced. Second, the sinker decreases the base resistance, thereby increasing the bipolar base current. Hence, the current limitation exhibited by structures containing buried layer and auxiliary emitter is removed.

The main process steps in IGBT fabrication are: epitaxial growth, buried layer using boron implantation, P^+ and N^+ sinker diffusion, polysilicon deposition, doping and photolithography, N^+-buffer and P-base implantations followed by drive-in, N^+-emitter implantation, shallow P^+ diffusion of collector and auxiliary emitter, contact window opening, Al metallization, and passivation.

Static characteristics of the LIGT have been experimentally measured over the temperature range 300–423 K. The latching current density of this LIGBT is 160 A/cm^2 at 423 K. The same is approximately 40 A/cm^2 for the conventional LIGBT. This LIGBT showed comparatively less reduction in latching current density with rise of temperature than did the conventional case. Dynamic latching has also been studied. About 0% decrease in maximum controllable current under inductive load condition has been measured.

Numerical simulation analysis has revealed that the latching mechanisms of the conventional LIGBT structure and the above structure are different. In the former, latching is caused by holes flowing through the P-base region and occurs in proximity of the channel. In the latter, latching takes place in the plane zone of the N$^+$-emitter/P-base junction where the P$^+$ sinker comes up with the P-base region. Thyristor triggering is initiated by holes moving through the buried layer and the P$^+$ sinker because the N$^+$ sinker diverts the holes away from the P base.

10.7 SELF-ALIGNED SIDEWALL-IMPLANTED N$^+$-EMITTER LATERAL IGBT (SI-LIGBT) WITH HIGH LATCHUP CURRENT CAPABILITY

In this structure, shown in Fig. 10.7, a small N$^+$ emitter is formed by the self-aligned phosphorus implantation through the sidewall of the trench after diffusing the P^{++} region at the bottom of the trench [6]. The P^{++} region is extended underneath the N$^+$ emitter, thereby drastically reducing the P-base resistance responsible for latching. Main process steps for SI-LIGBT fabrication are: P-base diffusion, self-aligned reactive ion etching for trench excavation in the center of the P base, boron implantation and

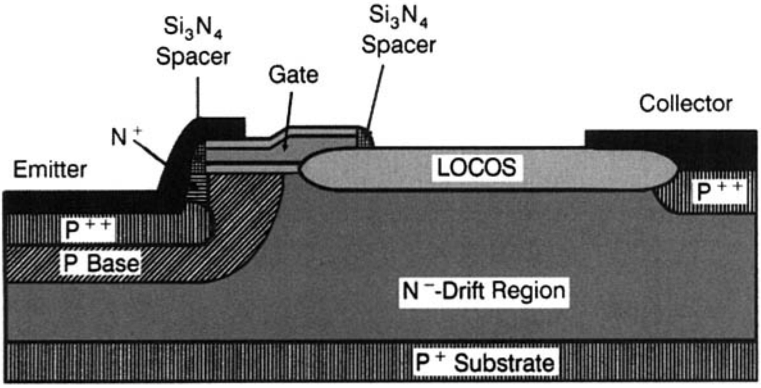


Figure 10.7 Self-aligned sidewall-implanted N$^+$-emitter lateral IGBT (SI-IGBT) for high latchup capability.

drive-in for the P^{++} region above the P base and P^{++} collector, and self-aligned 45° tilted implantation of phosphorus through the trench sidewall for producing the tiny N^+ emitter.

The SI-LIGBT does not require any additional photomask for the formation of P^{++} region and P base due to the self-aligned process. Hence, the SI-LIGBT area is smaller than in the traditional approach. The room-temperature latching current of the SI-LIGBT has been recorded as 1550 A/cm^2 which remains above 1000 A/cm^2 at 125°C, thus verifying the advantage of the sidewall implanted emitter.

10.8 IMPROVED LIGBT STRUCTURE FOR LARGER FBSOA

The lateral IGBTs suffer from latchup due to the parasitic PNPN thyristor between the collector and emitter. As a result, the forward bias safe operating area (FBSOA) of the device is impaired. This is especially applicable to those devices fabricated by the CMOS/BiCMOS technology, where the emitter doping concentration is restricted by the low threshold voltage necessary for the CMOS circuit. But SOA determines the short-circuit capability of IGBT. Therefore, a larger SOA is essential. Significant improvement in the forward bias safe operating area of a lateral IGBT (LIGBT) is achieved by the incorporation of a shallow, lightly doped P-layer in the drift region of a high-voltage lateral IGBT (Fig. 10.8) [7]. The basic cause of the

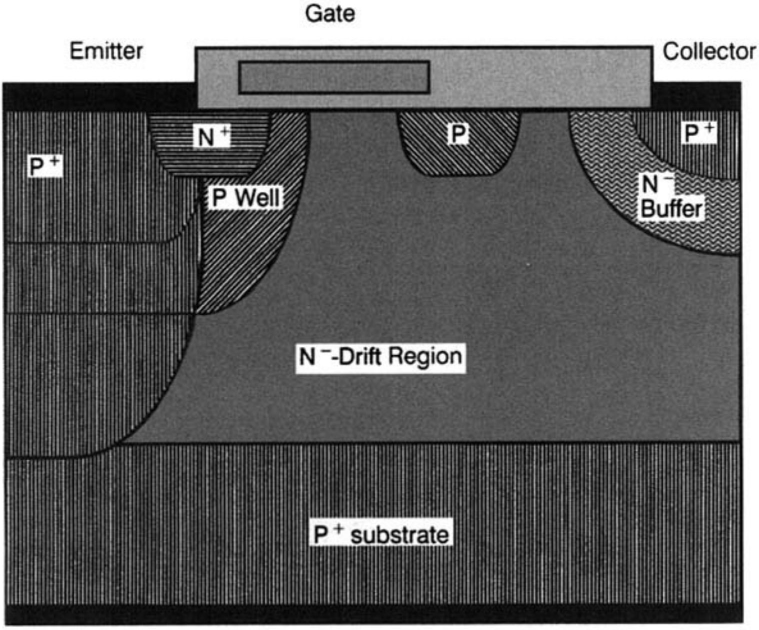


Figure 10.8 Modified lateral IGBT structure with a shallow P layer under the gate extension.

improvement, is the surface electric field reduction in the ON state. However, the length of the P layer included in the device has no derogatory influence on the conduction characteristics of the IGBT, nor does it increase the complexity of the HV-CMOS/BiCMOS process.

10.9 LATERAL IGBT WITH INTEGRATED CURRENT SENSOR

Current sensing [8] is required in power electronic systems for two purposes, namely, the closed-loop control application and as a protective measure to detect when the current crosses the safety limit so that current magnitude is restricted. At the same time, a mandatory requirement is that the normal device operation should not be impaired. Hence, the current sensor needs to draw a small current, but not so small that noise interference vitiates the measurements. The current to be sensed is the lateral current component—that is, the emitter current. Generally, the emitter current equals the collector current, but there may be a small difference due to the substrate leakage current. For dielectrically isolated devices, this difference is imperceptibly small.

The emitter current comprises two carrier components, namely, the hole and electron currents. The hole current mainly traverses the P-base region, while the electron component flows through the channel to the N$^+$ emitter. The electron current flowing through the P base is very small. The ratio of hole current to electron current is called the *current composition ratio*. This ratio varies with current density and gate voltage. Hence, both components must be accurately measured to obtain a realistic estimate of the emitter current.

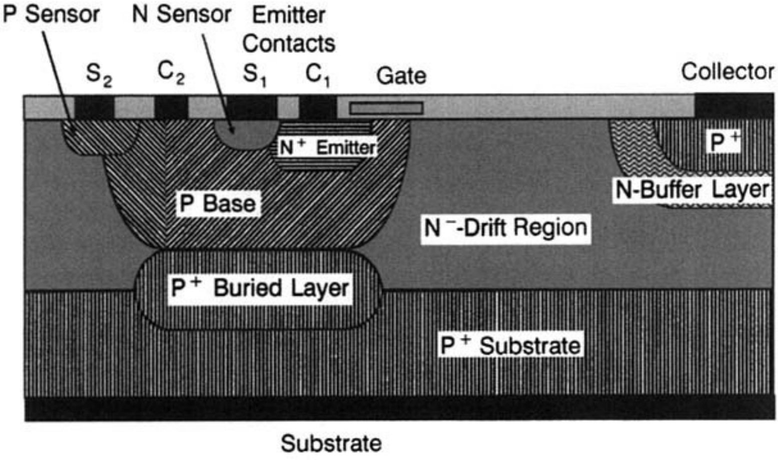


Figure 10.9 Integration of the current sensor in the LIGBT structure.

Figure 10.9 presents the structure of the current sensor. There are two sensor contacts S_1 and S_2. The sensor contact S_1 is located on an N layer, beside the emitter, to detect a portion of the electron current. The sensor contact S_2 is placed on a P layer, formed by a shallow diffusion into the P base, to divert the small amount of hole current to the contact S_2. The doping profiles of the N sensor and the P sensor are first defined. The sensor structure is then optimized with reference to the contact size and position and the doping profiles to reduce the variation in sensing ratio so that the (electron current at S_1/electron current at C_1) = (hole current at S_2/hole current at C_2). Thus the variation of lateral current sensing ratio was confined within a narrow range of 0.95–1.1 mA/A from 250 to 450 K, over a broad range of current density (more than 1400 A/cm^2) and for gate voltages up to 70 V and during transient turn-off.

10.10 DIELECTRICALLY ISOLATED FAST LIGBTs

In the dielectrically isolated LIGBT (DI-LIGBT), for smart power ICs [9], the use of electron irradiation for turn-off time control is avoided because of its adverse influence on other circuits present on the chip. Two viable alternatives include (a) the collector-shorted LIGBT (CS-LIGBT) (Fig. 10.10), in which the N$^+$ short is linearly placed behind the P$^+$ collector, and (b) the segmented collector (SC-LIGBT) structure, in which the N$^+$ short is located, along the device width in the third dimension followed by P$^+$ collector, and this sequence is successively repeated. The notable advantage

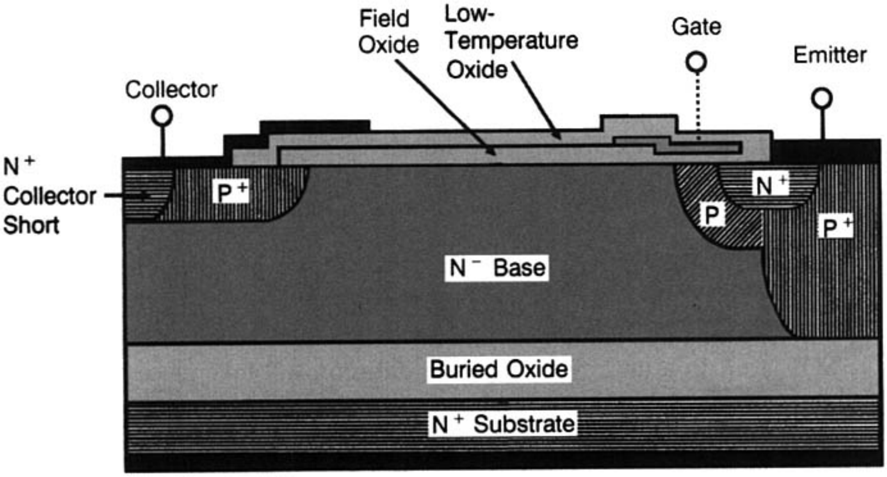


Figure 10.10 DI-collector-shorted IGBT.

is that the variation of P^+-collector segment width during device layout can control the trade-off between forward drop and turn-off time without any need to change the collector finger width and, therefore, the IGBT chip area. The DI CS-LIGBT showed a comparable forward drop to the conventional structure because the carrier transport at high current densities is mainly controlled by the drift component and the collector shorts do not greatly alter the electron–hole concentrations. Also, the DI CS-LIGBT was found to exhibit lower forward conduction drop than the DI SC-LIGBT. The higher ON-state voltage drop of the SC-LIGBT was the result of reduced injection of carriers from the P^+ collector, not only due to presence of N^+ shorts on the collector periphery, which reduce the effective collector region, but also because a large portion of the collector periphery is unable to inject carriers, since the forward bias developed by the electrons flowing under the P^+ collector into the shorts is inadequate to initiate injection. The turn-off behavior of both the CS-LIGBT and SC-LIGBT was superior to the conventional IGBT, because they were faster in operation. But the SC-LIGBTs were found to latch at lower current densities primarily due to nonuniformity of current flowing in the N^--base and P-base regions.

10.11 LATERAL IGBT IN THIN SILICON-ON-INSULATOR (SOI) SUBSTRATE

True dielectric isolation is furnished by SOI substrates. But generally, substrates of thickness more than 10 μm are used. Both the isolation of devices and the integration of logic components are greatly simplified by using ultrathin substrates less than 1 μm thick [10]. For supporting the breakdown voltage, a linearly graded doping profile is employed in the drift region, for achieving a uniform lateral electric field. By this approach, a high breakdown voltage is obtained along with a low forward voltage drop and small turn-off time. This provides the desirable electrical characteristics of high-speed power integrated circuits (PICs). A critical parameter is the SOI substrate thickness. Too thin a substrate increases the carrier recombination at its top and bottom interfaces, inhibiting conductivity modulation and thereby increasing the forward potential drop of the IGBT. Conversely, too thick a substrate poses problems in the formation of a linearly graded profile. This profile is realized by ion implantation, through a series of openings in a mask. The openings are large in size near the collector and smaller near the emitter, so that different amounts of impurities are introduced at these regions. The ion implantation is followed by lateral diffusion of impurities, in the SOI layer. In a thick SOI layer, the diffusion will take place laterally as well as vertically, so that the resulting profile is not linear. A high breakdown voltage is obtained, only by using an impractically large buried oxide thickness. Breakdown voltage, up to 720 V, was obtained in the optimum SOI layer thickness of 0.5 μm, with the N-buffer layer concentration of 1×10^{18} cm^{-3} and buried oxide thickness of 4 μm, giving the best trade-off between

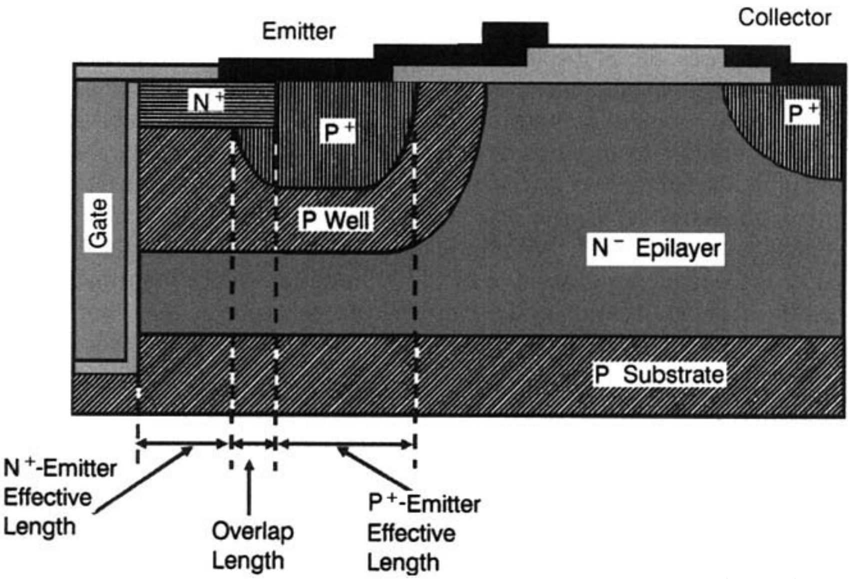


Figure 10.11 Cross-sectional view of the lateral trench-gate bipolar transistor (LTGBT).

breakdown voltage and forward drop. The forward drop was 6 V at 100 A/cm^2 and at a turn-off time of 140 nsec, in the same device. The device operation is free from latching up to 200 A/cm^2 at 25°C.

10.12 LATERAL TRENCH-GATE BIPOLAR TRANSISTOR (LTGBT) FOR IMPROVED LATCHUP CHARACTERISTICS

This structure [11] differs from a conventional LIGBT in the location of the MOS gate region and the N$^+$/P$^+$ emitter regions (Fig. 10.11). It comprises a trench gate with the positions of the emitter and the channel interchanged, with respect to the conventional LIGBT. A vertical trench forms the MOS gate, and the connection of P well with N$^+$ emitter is made through P$^+$ emitter. In the forward conduction mode of operation, a negative voltage is applied to the emitter with respect to the collector. As in the conventional LIGBT, collector current begins to flow at a gate voltage higher than the threshold voltage and a collector voltage greater than one PN-junction potential drop. With increasing collector voltage, the collector PN junction injects a large number of holes into the N$^-$ epilayer. A fraction of these holes undergo recombination with the electrons injected by the vertical channel. The remaining holes flow from the N$^-$ epilayer to the P well. The P$^+$ emitter gathers these holes, without allowing said holes to flow through the region underneath the N$^+$ emitter. Thus the existence of the P well between the channel and the collector regions facilitates the flow of holes toward the P$^+$

emitter and reduces the tendency of hole flow in the region below the N^+ emitter, thereby preventing latching: for example, an N^+-emitter length of 5 μm increases the static latching current density by 2.3 times the LIGBT, whereas the dynamic latching current density was increased by 4.2 times.

Further latchup immunization is accomplished by shortening the length of the N^+-emitter region. Practically, the emitter length shortening is restricted by photolithographic constraints. Without the help of complex photolithography, the alternative course is to overlap the N^+- and P^+-emitter regions. The junction depth of the P^+ emitter is made greater than that of the N^+ emitter, with the dopant concentration of the N^+ emitter higher than that of the P^+ emitter. Thus, on the surface, the N^+-emitter region is long, but its role in latching is suppressed by the portion of the P^+ emitter below the N^+ emitter. Therefore, depending on amount of overlapping, an effective N^+-emitter length is defined, which is less than the total N^+-emitter length, by the overlap length. Then, the effective N^+-emitter length determines the maximum latchup current. By progressively increasing the overlapping between the N^+- and P^+-emitter diffusions, the effective N^+-emitter length was made comparatively small; and at a length of 2 μm, no latching was observed in the LTGBT. However, this artifice increased the threshold voltage by 0.8 V.

Conductivity modulation in the LTGBT occurs not only in the N^--drift region, but also deeper into the N^- epilayer and the P substrate. At lower collector voltages (< 2.6 V), a large resistance is offered to electron flow. This is because the electrons injected from the vertical channel have to traverse a long path length in the LTGBT, deep into the bulk, before recombination. The large resistance raises the forward drop of LTGBT above the LIGBT; but as the collector voltage increases, significant hole injection from the collector side is initiated. Then the electron injection from the vertical channel into the N^--epilayer and P-substrate regions, combined with the hole injection from the collector, produces enhanced conductivity modulation. In this voltage range, the conductivity modulation effect in LTGBT is stronger than that in LIGBT. Hence, the forward conduction characteristics of the LTGBT are superior to the LIGBT.

10.13 TRENCH PLANAR INSULATED GATE BIPOLAR TRANSISTOR (TPIGBT)

This device [12] is an improvement over the planar fine lithography IGBT (FL-IGBT). In the FL-IGBT, fine-resolution photolithographic techniques are utilized to obtain a low ON-state voltage. In the TPIGBT (Fig. 10.12), incorporation of a trench gate between the emitter cells of the FL-IGBT further enhances its performance, by reducing the IGBT forward voltage drop without affecting its breakdown voltage capability. As seen in Section 10.2, the IGBT is a combination of PNP transistor with PIN rectifier. It has been shown [13] that the forward conduction characteristics of the IGBT are

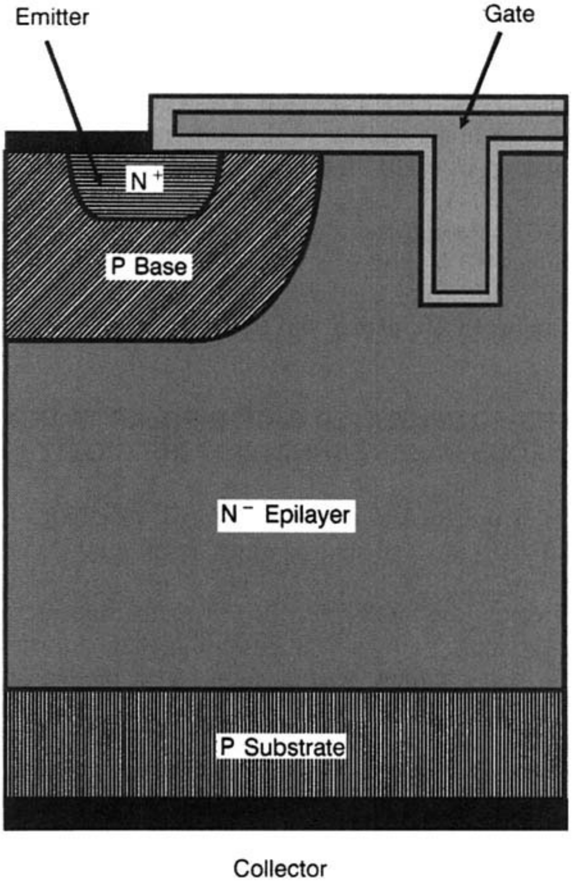


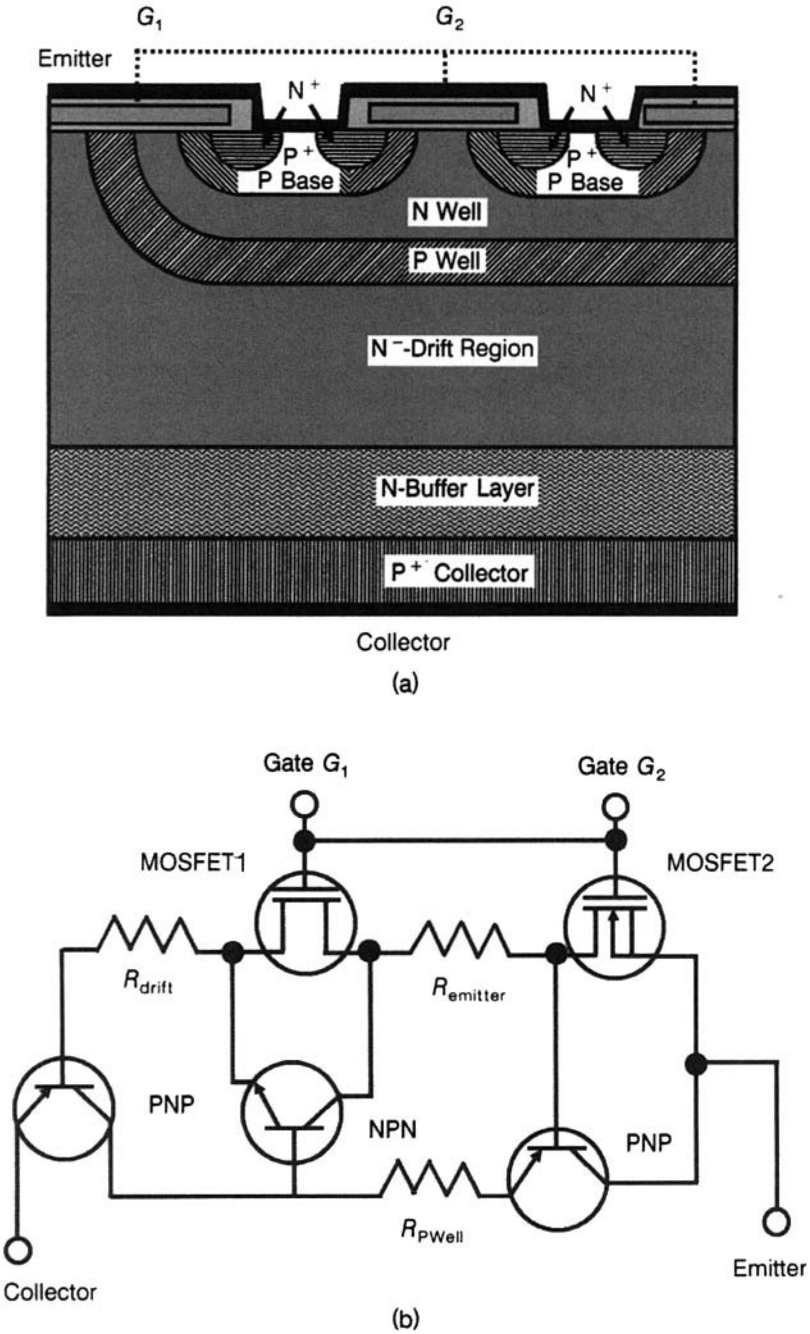
Figure 10.12 Schematic cross section of the trench planar insulated gate bipolar transistor.

improved by reducing the contribution of the PNP transistor in favor of the PIN rectifier. If the PIN rectifier effect is augmented near the emitter region, the saturation characteristics are degraded. Therefore the PIN rectifier effect must be transferred away from the main emitter region to obtain a good, forward-bias, safe operating area (FBSOA). In the TPIGBT, this is accomplished by forming a large accumulation layer along the trenches inserted between the emitter cells. Thus the TPIGBT achieves injection enhancement. The TPIGBT utilizes another property in that the conductivity modulation of the drift region is increased by squeezing the flow of holes and allowing these carriers to accumulate in the region below the emitter. In the TPIGBT, the gate extends over the channel region of the P well, a short distance beyond the P well, up to a trench filled with N-type polysilicon, over an oxide film. The N^--drift region, separating the P well and the trench, is referred to as the *mesa region*. By using a narrow mesa region of width less than 3 μm, the hole flow is constricted. Furthermore, with the trench depth being less than that of the P^+-isolation region, premature electric field crowding at the corners of the trench is avoided. Thus a high electric field in the MOS channel region cannot develop. This permits the use of a thin gate oxide, ~ 250 Å, which is a quarter of the conventional gate oxide thickness. The ON-state voltage of a TPIGBT, having a trench depth of 3 μm, and gate oxide thickness of 250 Å, is less by 0.5 V at 100 A/cm^2 current density, than that of a FL-IGBT. Upon increasing the gate oxide thickness to 1000 Å, this difference becomes 0.3 V. Use of a smaller trench depth further diminishes the forward drop. Regarding the switching losses, the TPIGBT represents a good trade-off between the planar and trench IGBTs.

10.14 CLUSTERED INSULATED GATE BIPOLAR TRANSISTOR IN HOMOGENEOUS BASE TECHNOLOGY (HB-CIGBT)

CIGBT [14, 15], is designed to extend the IGBT capabilities to the blocking capability range of 6 kV and above, which is presently dominated by the GTO and IGCT.

The CIGBT employs a MOS-controlled thyristor concept, for low forward conduction losses, along with current saturation characteristics, at high gate bias, through a "self-clamping phenomenon," at a predetermined collector voltage. It is formed by the periodic repletion of unit cells called the *cluster cells*, consisting of a number of closely packed IGBT emitter cells, confined in a floating N well and P well. In the CIGBT half-cell, known as the *cluster half-cell* (Fig. 10.13a), gate 1 (G_1) is the turn-on gate while gate 2 (G_2) is the control gate. The gates G_1 and G_2, are connected together. Similarly, the emitter contacts are also connected together and grounded. The surface concentration of the P well is lower than that of P base. The threshold voltage of G_2, being higher than that of G_1, determines the threshold voltage of the IGBT. For ON-state operation, a positive potential is applied to the P collector, and a positive voltage greater than the threshold voltage is applied to the gate G_2. This initiates the flow of electron current into the N^--drift region through the gate G_1. During this operation, the N well is maintained at ground potential via the channel under gate G_2, connected to the grounded emitter contact. Also, the P well is in a floating condition and its voltage rises, by capacitive coupling, to the applied collector bias. As soon as the potential drop across the N-well/P-well junction exceeds the built-in potential across the junction, carriers are injected across this junction and thyristor action starts across the structure of the P^+-collector/N-buffer layer, N^--drift region/P well/N well. Then, the gate G_1 loses control. The potentials of the N well and P well increase due to conductivity modulation, but across the reverse-biased P-base/N-well junction, the depletion region stretches to the P well at a certain collector voltage, up to the punchthrough limit.


G_1
G_2
Emitter
N^+
N^+
P^+
P Base
P^+
P Base
N Well
P Well
N^--Drift Region
N-Buffer Layer
P^+ Collector
Collector
Gate G_1
Gate G_2
MOSFET1
MOSFET2
R_{drift}
$R_{emitter}$
PNP
NPN
PNP
R_{PWell}
Collector
Emitter

(a)

(b)

Figure 10.13 Cross-sectional view and equivalent circuit model of CIGBT. (a) Half-cell of CIGBT. (b) Equivalent circuit representation of CIGBT.

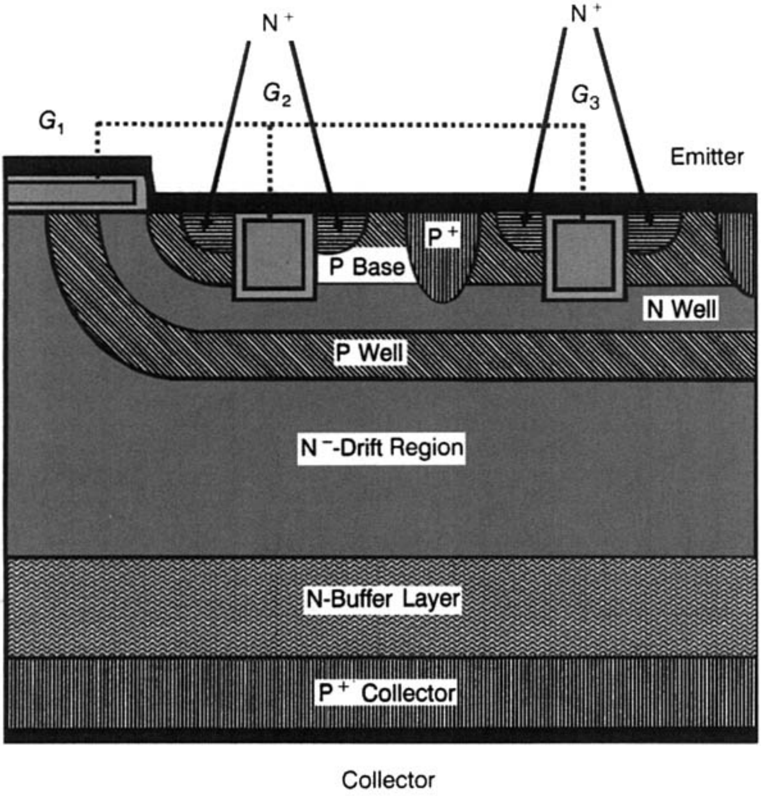


Figure 10.14 Half-cell of trench clustered CIGBT (TCIGBT).

Consequently, there is a self-potential clamping of the N-well and P-well regions. Thus the current saturates at high gate voltage. Figure 10.13b presents the electrical equivalent circuit of CIGBT.

10.15 TRENCH CLUSTERED INSULATED GATE BIPOLAR TRANSISTOR (TCIGBT)

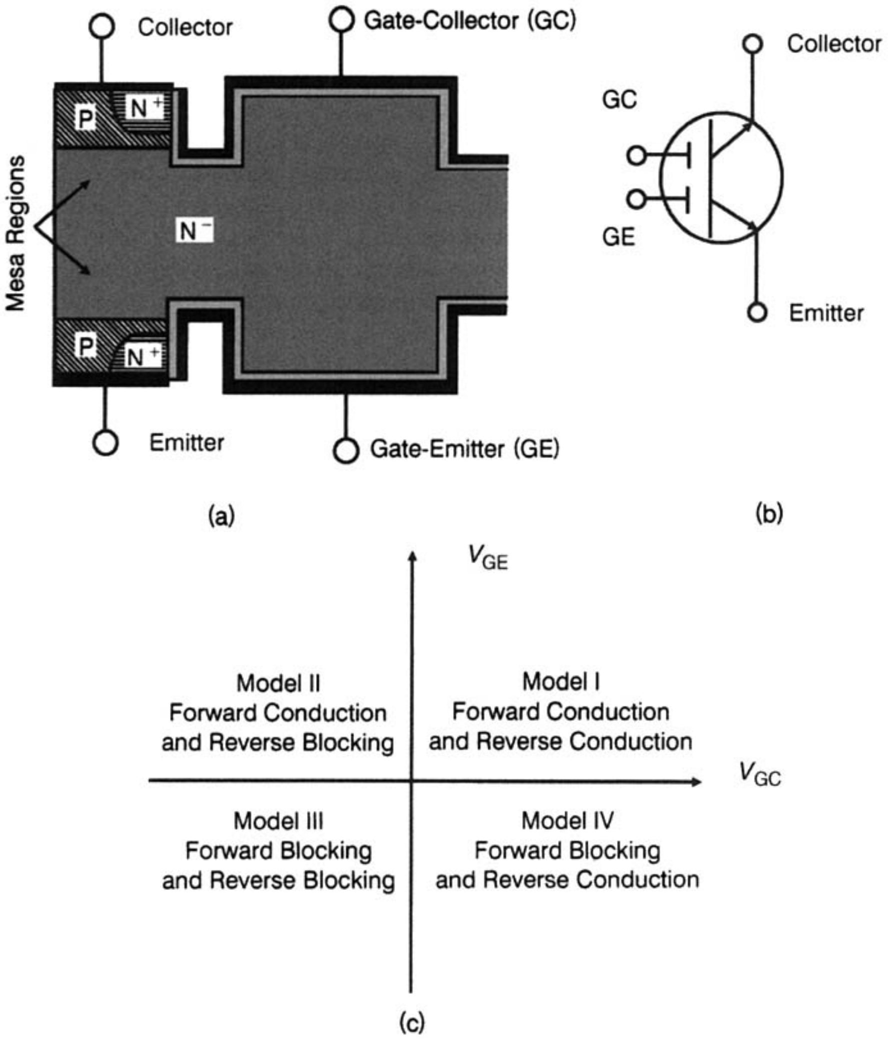
This structure [16], shown in Fig. 10.14, employs shallow trenches, protected from high collector potential by a self-clamping technique. It contains ~ 5–10 trench-gated emitter cells, which are surrounded by a common N well and a P well. The planar MOS gate G_1 located over the P well, and connected to the trench gates, is used for turning on the device but does not play any subsequent role in device operation. The structure is comprised of two PNP transistors and one NPN transistor. The two PNP transistors are (a) PNP1, consisting of the P^+ collector substrate, the N^--drift region, and the P well, and (b) PNP2, consisting of the P well, the N well, and the P base. The NPN transistor is composed of the N^--drift region, the P well and the N well. The transistors PNP1 and NPN comprise the main thyristor.

As soon as the N-channel gates are in the ON state, grounding of the N well occurs through the inversion and accumulation layers formed. But the P well is in a floating condition, so its potential rises with the collector potential. Eventually, when the voltage drop across the P-well/N-well junction exceeds the built-in potential of the diode, the NPN transistor is turned on triggering the main thyristor into action. Then, the N well and P well potentials increase continuously, with collector potential, but the depletion region of the reverse-biased P-base/N-well junction extends downwards, causing punchthrough of the transistor PNP2. Upon further rise of collector potential, the voltage is dropped across the P-well and N^--drift regions. Consequently, current saturation occurs and self-clamping is accomplished. For adjustment of the self-clamping voltage, the depth of the P^+ region is altered. Saturation current of the TCIGBT varies accordingly.

The TCIGBT provides a lower forward potential drop than the TIGBT due to thyristor action. Simulation studies have shown that the ON-state potential drop of the 2.4-μm, 1200-V NPT-TCIGBT is 1.4 V at 100 A/cm^2 as compared to 1.9 V for the TIGBT at the same current density, indicating a 25% reduction of ON-state losses. Similarly, the turn-off losses decreased by 28%. It also has a positive temperature coefficient, of forward voltage, thereby facilitating paralleling of devices. Saturation current of the TCIGBT is lower than that of the TIGBT. Thus, the TCIGBT is a superior substitute for the TIGBT and also for the GTO, in the high power sector.

10.16 DOUBLE-GATE INJECTION-ENHANCED GATE TRANSISTOR (DG-IEGT)

This is an improved IEGT structure [17], which not only eliminates the trade-off between ON-state loss and OFF-state loss but also has bidirectional conduction and blocking capabilities. The schematic cross section of DG-IEGT is shown in Fig. 10.15a, and its circuit diagram symbol is shown in Fig. 10.15b. The absence of a buffer layer, on the collector side, is observed. Hence, it is a symmetrical structure, which conducts currents and blocks voltages of both polarities. Only the mesa regions on the extreme left are connected to the emitter contact. Referring to Fig. 10.15c, the device has four possible modes of operation: (i) **$V_{GE} > 0$, $V_{GC} > 0$:** Here, V_{GE} is the potential difference between gate and emitter, at the emitter. This mode is characterized by *bidirectional conduction.* It has two steady states, namely, forward conduction and reverse blocking states. The conduction takes place like a MOSFET, without conductivity modulation. This mode results in higher conduction losses and is to be avoided. (ii) **$V_{GE} > 0$, but $V_{GC} < 0$:** The two steady states in this mode are the forward conduction and reverse


Collector
Gate-Collector (GC)
Mesa Regions
P
N+
N−
P
N+
Emitter
Gate-Emitter (GE)
GC
GE
Collector
Emitter
V_{GE}
Model II
Forward Conduction
and Reverse Blocking
Model I
Forward Conduction
and Reverse Conduction
V_{GC}
Model III
Forward Blocking
and Reverse Blocking
Model IV
Forward Blocking
and Reverse Conduction

(a)

(b)

(c)

Figure 10.15 Structural view and operational modes of DG-IEGT. (a) DG-IEGT structure. (b) Circuit symbol of DG-IEGT. (c) Four operational modes of DG-IEGT.

blocking. In the forward conduction mode, the voltage drop is extremely low because of minority carrier injection and the conductivity modulation phenomenon, as in a P-I-N rectifier. Due to superior trade-off between the forward conduction and blocking capabilities, this mode lends itself to widespread application. (iii) **$V_{GE} < 0$, but $V_{GC} < 0$:** This is a bidirectional blocking mode, which is useful for blocking an AC signal, without applying any control bias to the gate terminals. (iv) **$V_{GE} < 0$, but $V_{GC} > 0$:** This mode is similar to mode (ii), with the reversal of current and voltage directions.

Thus the DG-IEGT is a four-quadrant switch that will find applications for replacing multiple switches and diodes, greatly simplifying system design.

10.17 SiC IGBTs

Due to the unique properties of silicon carbide in the harsh environments of high-power, high-temperature, and high-frequency electronics, there is a recent resurgence of interest in this compound, because high-quality bulk and custom epitaxial wafers have become available. But, among the 150 polytypes, only three varieties—the 3C-SiC, 6H-SiC, and 4H-SiC—are most common; and among these three the 4H-SiC polytype is chosen, due to the isotropic nature of its properties [18, 19]. Furthermore, the electron mobility in 4H-SiC is double of that in 6H-SiC, perpendicular to the *c* axis, and almost a decade more than that of 6H-SiC, parallel to the *c* axis. A primary process advantage of SiC (among the compound semiconductors GaAs and GaN) is that it can be thermally oxidized to give SiO_2. Demonstration of MOS structures, with high breakdown strength and moderate density of interface states, has enabled the fabrication of MOS-based devices and IGBTs with this material. As in silicon, the electron mobility in SiC is higher than hole mobility, and so the transconductance of N-channel SiC MOSFET is larger than that of the P-channel variant, while the current gain of the active PNP transistor in the N-channel IGBT is lower than that of the active NPN transistor in the P-channel IGBT. Hence, the forward drops of N-channel and P-channel IGBTs are nearly equal, and their performance is competitive. However, in both 4H-SiC and 6H-SiC polytypes, the impact ionization coefficient of the holes is larger than that of electrons, implying that the number of electron–hole pairs, produced by a hole traversing a unit distance (1 cm) through the depletion region in the direction of the electric field is more than the number of electron–hole pairs generated by an electron. Thus the active NPN transistor of a P-channel IGBT is more rugged than the active PNP transistor of N-channel IGBT, and hence the safe operating area of a P-channel IGBT is larger than that of an N-channel IGBT. For this reason, the P-channel IGBT is preferred over the N-channel IGBT, in both the SiC polytypes.

The SiC IGBT has taken a back seat to SiC power MOSFET research and development, while SiC IGBTs continue to be a relatively unexplored area of SiC power semiconductor research [20]. In the N-channel trench 6H-SiC IGBT [21], the electron mobility in the inversion layer is appreciably small, < 1 cm^2/V-sec. Hence the conductivity modulation of drift region is less and therefore the forward current is low, $\sim$ mA. This IGBT has breakdown voltage of ~ 300 V. A 200-V breakdown voltage, P-channel UMOS IGBT [22], fabricated in 4H-SiC, carries a higher current (> 10 mA) than does the 6H-SiC IGBT. Developmental efforts made for SiC MOSFETs and IGBTs have shown that these devices can reduce the energy consumption in power electronic systems. Speculation has been made that SiC IGBTs may be made with 25-kV blocking voltage capability and superior switching performance.

It is noteworthy that in SiC, a smaller drift layer thickness and higher background doping is used for a given breakdown voltage. Hence the IGBT turn-off times are smaller in SiC than in Si. As the lifetime in SiC decreases from 25 μsec to 5 μsec, the simulated turn-off time falls from 140 nsec to 50 nsec.

Extensive simulation studies have been performed of SiC power devices including trench gate IGBTs in the voltage range of 1–10 kV [23]. For voltages below 4 kV, the drift region thickness is very small, making punchthrough IGBTs unnecessary. In 4H-SiC IGBTs, the built-in potential of the P^+-substrate/N^--base junction (2.7 V) must be exceeded for current conduction. So the forward conduction characteristics of SiC IGBTs are not as favorable as those of SiC MOSFETs and BJTs. A comparison has been made between 3.3-kV Si IGBT and 4-kV SiC MOSFET. Whereas the MOSFET can handle a current density of 130 A/cm^2, the current handling capability of Si IGBT is 26.7 A/cm^2. Another noteworthy inference from these simulations is that the degradation of the current rating of a SiC IGBT with temperature is not as severe as that of SiC MOSFETs, or bipolar transistors. For instance, at a current density of 200 A/cm^2, the forward voltage drop of a 1-kV SiC IGBT is 4.29 V at 300 K and 3.9 V at 600 K. For the 10-kV SiC IGBT, the forward drops at 300 K and 600 K are 3.9 V and 4.6 V, respectively. So, thermal effects do not hinder the current sharing of IGBTs. The 7-kV SiC IGBT has a greater current carrying capability (223 A/cm^2 at 4 V) than does the Si thyristor. So, the SiC IGBT has the potential to replace the Si thyristor.

Above 3 kV, SiC IGBTs along with GTOs and SIThs (Static Induction Thyristors) are competing devices, but the requirement of a turn-off snubber and slower switching speed makes the IGBTs and SIThs more attractive than GTOs. So above 3 kV, SiC IGBTs and GTOs are comparable in their capabilities, both in current-carrying capacity and switching characteristics. Now comparing the SiC IGBTs and SIThs, both at room temperature, at a frequency of 1 kHz, the SiC SIThs dominate over the IGBTs. But at 600 K and a frequency of 10 kHz, SiC IGBTs are favored over the SITh devices. Thus the IGBTs exhibit a distinct superiority over the SITh in the high-temperature and high-frequency range. In addition, SIThs must be provided a large gate current during turn-off and are prone to turn-off failure due to gate debiasing, making the IGBTs still more attractive.

It is envisioned that in silicon carbide power devices, a strong competitor of the Si IGBT will be the MOS turn-off thyristor (MTO) [24]. This promising SiC device is a hybrid combination of MOSFET with gate turn-off thyristor (GTO). Due to its high-power capability, easy turn-off capability, up to 500°C operation, and resulting reduced cooling requirements, it is envisaged that a SiC MTO with an antiparallel diode will be an effective substitute for an Si IGBT/diode, for construction of an inverter module at high-temperature, high-power DC motor controls. It is further predicted that SiC MOSFETs and IGBTs will be unreliable at elevated temperatures due to the fundamental problems of charge injection and trapping in dielectric layers. At present the interface trap density ($\sim 5 \times 10^{11}$ cm^{-2}) and fixed oxide charge density ($\sim 1 \times 10^{12}$ cm^{-2}) in gate oxides, grown by thermal oxidation of SiC, are very high, so that limited success has been achieved with MOS-based SiC devices.

10.18 SUMMARY AND TRENDS

Power electronics has been in a dynamic state over the past two decades fueled by the developments in power semiconductor devices and modules. In the low- and medium-power range (1–10 kVA), FET-driven devices (MOSFETs and IGBTs) will dominate the market, with modules used, wherever specifications are met. Applications include automotive electronics, small motors, and domestic appliances. At the high-power end (several kVA to > 1 MVA), 6.5 kV and beyond, such as heavy-duty traction drives and industrial electronics, GTOs, IGCT, and IGBT modules will be strong contenders. An optimized trench IGBT derives the advantage of an enhanced PIN diode effect, lowering the ON resistance without compromising with the turn-off behavior. Cell size reduction, elimination of the parasitic JFET effect, latching immunity, and enhanced carrier injection on the emitter side has brought considerable improvement in device characteristics. The trench IGBT is far superior to its DMOS counterpart because, for DMOS IGBTs with thick N-drift region, the carrier modulation on the emitter side is appreciably weak. So, the trench IGBT will be a key player in the medium to high and ultra-high voltage ranges and has the potential to be a substitute for GTO thyristors at high voltages. Presently, GTOs and IGCT modules are in intense competition with IGBT modules for the high current and high-voltage applications. GTO and IGCT are superior, because they are based on thyristor action and so have a lower forward drop. However, these devices suffer from complex drive circuit requirements, so that extension of IGBT to this rating is promising for the replacement of GTO and IGCT by IGBT.

REVIEW EXERCISES

10.1 Define the "parallel effect" and the "coupled effect" in the PIN diode–PNP transistor model of trench IGBT.

10.2 What is the design aspect ratio of TIGBT. Define the term "optimized TIGBT structure." How does a high design aspect ratio lead to enhanced conductivity modulation on the emitter side without any degradation of switching behavior?

10.3 Elucidate the concept of dynamic buffer IGBT (DB-IGBT). Draw the cross-sectional diagram of the collector-side gate for broad and narrow mesa regions and sketch the corresponding energy band diagrams. How does the buffer zone behave under the influence of a positive and negative gate potential?

10.4 Why is the reverse blocking capability of an IGBT sacrificed in favor of forward conduction and blocking abilities? What is the disadvantage of using a series-connected diode? Suggest a lateral IGBT configuration that provides both forward and reverse blocking capabilities.

10.5 Why the highly doped sinker region included in the lateral IGBT for high-temperature ruggedness does not impair the blocking capability of the device during the OFF state? Explain the two ways by which it enhances the performance of the IGBT in the ON state. How does the latching mechanism of this IGBT differ from a conventional lateral IGBT?

10.6 How is the tiny N^+ emitter formed in an SI-IGBT? How is the P-base resistance reduced in this structure to avert latching?

10.7 What is the reason for the poor forward bias safe operating area of a lateral IGBT? What is the impact of including a shallow P-layer on SOA?

10.8 Why electron irradiation is not a suitable technique for lifetime killing in DI-LIGBT used in a smart power IC? Describe the location of N^+ collector short in CS-LIGBT and SC-IGBT. Which of these LIGBT structures gives a lower forward drop and why?

10.9 What do you understand by the acronyms: LTGBT and TPIGBT?

10.10 How does the incorporation of a trench gate along with interchanging the positions of the channel and emitter in conventional IGBT help in improving the latching current density? Explain the operation of LTGBT with a diagram.

10.11 In which forward voltage range are the electrical characteristics of an LTGBT inferior to the LIGBT, and in which range they are superior? Why?

10.12 Define the effective N^+ emitter length in LTGBT. How does the overlapping of N^+ and P^+ emitter regions with a differential junction depth between them, help in obtaining a short N^+ emitter without expensive photolithography?

10.13 How does the incorporation of a trench gate between emitter cells result in a significant reduction in the forward voltage drop in a TPIGBT?

10.14 How is premature electric field crowding at the trench corners avoided in a TPIGBT? How is the PIN rectifier effect enhanced in a TPIGBT?

10.15 How is a low forward voltage drop achieved in a clustered IGBT? Explain the "self-clamping effect" in a CIGBT and the current saturation at high gate bias?

10.16 Draw the cross-sectional diagram and explain the operation of a TCIGBT. Outline its advantages over the trench gate IGBT.

10.17 What are the four operational modes in a DG-IGBT? Which of these modes are preferred, and which must be avoided for proper device function? Explain with arguments. How will the DG-IGBT concept help in system design simplification?

10.18 How many polytypes of SiC are known? Which polytypes of SiC are popular? Among 4H-SiC and 6H-SiC, which has a higher electron mobility? Which exhibits isotropic behavior of properties? Mention one process advantage that makes SiC amenable to MOS device technology.

10.19 Compare the silicon carbide N-channel IGBT with the P-channel IGBT in terms of their transconductances and safe operating areas.

10.20 Highlight the advantages and disadvantages of SiC IGBTs, GTOs, and SITHs above 3 kV. For what type of applications are SiC IGBTs superior?

10.21 Name the devices vying with IGBT for high voltage applications ≥ 6.5 kV.

10.22 Amongst the two classes of trench IGBT, which structure gives lower forward drop and why?

REFERENCES

1. (a) F. Udrea and G. A. J. Amaratunga, Optimum Carrier Distribution of the IGBT, *Solid State Electron.*, Vol. 44, No. 9, September 2000, pp. 1573–1583.
   (b) F. Udrea and G. A. J. Amaratunga, An ON-state Analytical Model for the Trench Insulated Gate Bipolar Transistor (TIGBT), *Solid State Electron.*, Vol. 41, No. 8, August 1997, pp. 1111–1118.
2. N. Thapar and B. J. Baliga, An Experimental Investigation of the ON-state Performance of Trench IGBT Designs, *Solid State Electron.*, Vol. 42, No. 5, May 1998, pp. 771–776.
3. S. Huang, K. Sheng, F. Udrea and G. A. J. Amaratunga, A Dynamic N-Buffer Insulated Gate Bipolar Transistor, *Solid State Electron.*, Vol. 45, No. 1, January 2001, pp. 173–182.
4. M. Mehrotra and B. J. Baliga, Reverse Blocking Lateral MOS-Gated Switches for AC Power Control Applications, *Solid State Electron.*, Vol. 42, No. 4, April 1998, pp. 573–579.
5. M. Vellvehi, P. Godignon, D. Flores, J. Fernández, S. Hidalgo, J. Rebollo and J. Millán, A New Lateral IGBT for High Temperature Operation, *Solid State Electron.*, Vol. 41, No. 5, May 1997, pp. 739–747.
6. D. S. Byeon, B. H. Lee, D. Y. Kim, M. K. Han, Y. I. Choi, and C. M. Yun, A Lateral Insulated Gate Bipolar Transistor Employing the Self-Aligned Sidewall Implanted N^+ Source, *International Electron Devices Meeting (IEDM), Technical Digest*, Dec. 6–9, 1998, San Francisco, IEEE, New York, pp. 687–690.
7. S. Hardikar, G. Cao, Y. Xu, M. M. De Souza, and E. M. S Narayanan, A Local Charge Control Technique to Improve the Forward Bias Safe Operating Area of LIGBT, *Solid State Electron.*, Vol. 44, No. 7, July 2000, pp. 1213–1218.
8. Y. C. Liang, G. S. Samudra, and V. S. S. Hor, Design of Integrated Current Sensor for Lateral IGBT Power Devices, *IEEE Trans. Electron Devices*, Vol. 45, No. 7, July 1998, pp. 1614–1616.
9. R. Sunkavalli and B. J. Baliga, Comparison of High-Speed DI-LIGBT Structures, *Solid State Electron.*, Vol. 41, No. 12, December 1997, pp. 1953–1956.
10. Y. -K. Leung, A. K. Paul, J. D. Plummer, and S. S. Wong, Lateral IGBT in Thin SOI for High-Voltage, High-Speed Power IC, *IEEE Trans. Electron Devices*, Vol. 45, No. 10, October 1998, pp. 2251–2254.
11. J. Cai, J. K. O. Sin, P. K. T. Mok, W.-T. Ng and P. P. T. Lai, A New Lateral Trench-Gate Conductivity Modulated Power Transistor, *IEEE Trans. Electron Devices*, Vol. 46, No. 8, August 1999, pp. 1788–1793.

12. O. Spulber, E. M. S. Narayanan, S. Hardikar, M. M. De Souza, M. Sweet and J. V. S. C. Bose, A Novel Gate Geometry for the IGBT: The Trench Planar Insulated Gate Bipolar Transistor (TPIGBT), *IEEE Electron Device Lett.*, Vol. 20, No. 11, November 1999, pp. 580–582.
13. F. Udrea and G. A. J. Amaratunga, A Unified Analytical Model for the Carrier Dynamics in Trench Insulated Gate Bipolar Transistor, in *Proceedings International Symposium on Power Semiconductor Devices and ICs (ISPSD'95)*, 1995, Yokohama, pp. 190–196.
14. E. M. S. Narayanan, M. Sweet, O. Spulber, M. M. De Souza, and S. C. Bose, Clustered IGBT—A New Power Semiconductor Device, *Proceedings International Workshop on Physics of Semiconductor Devices (IWPSD'99)*, 1999, Allied Publishers, New Delhi, pp. 1307–1312.
15. N. Luther-King, M. Sweet, O. Spulber, K. Vershinin, C. K. Ngw, S. C. Bose, M. M. De Souza, and E. M. S. Narayanan, The 6.5 kV Clustered Insulated Gate Bipolar Transistor in Homogeneous Base Technology, *Solid State Electron.*, Vol. 45, No. 1, January 2001, pp. 71–78.
16. M. Sweet, K. Vershinin, C. K. Ngw, L. Ngwendson, J. V. S. C. Bose, M. M. De Souza, and E. M. S. Narayanan, A Novel Trench Clustered Insulated Gate Bipolar Transistor (TCIGBT), *IEEE Electron Device Lett.*, Vol. 21, No. 12, December 2000, pp. 613–615.
17. Y. Bai and A. Q. Huang, Comprehensive Investigations of High Voltage Non-Punchthrough Double Gate-Injection Enhanced Gate Transistor, *Solid State Electron.*, Vol. 44, No. 10, October 2000, pp. 1783–1787.
18. T. P. Chow, V. Khemka, J. Fedison, N. Ramungul, K. Matocha, Y. Tang, R. J. Gutmann, SiC and GaN Bipolar Power Devices, *Solid State Electron.*, Vol. 44, No. 2, February 2000, pp. 277–301.
19. E. Martin, J. Jiménez, and M. Chafai, Microraman Study of Crystallographic Defects in SiC Crystals, *Solid State Electron.*, Vol. 42, No.12, December 1998, pp. 2309–2314.
20. A. Elford and P. A. Mawby, Emerging Silicon Carbide Power Device Technologies, *J. Wide Bandgap Mater.*, Vol. 7, 2000, pp. 179–191.
21. N. Ramungul, T. P. Chow, M. Ghezzo, J. Kretchmer, and W. Hennessy, A Fully Planarized 6H-SiC UMOS Insulated Gate Bipolar Transistor, *54th Annual Device Research Conference Digest*, 1996, pp. 56–58.
22. R. Singh, S.-H. Ryu, and J. W. Palmour, High-Temperature, High-Current P-channel UMOS 4H-SiC IGBT, *57th Annual Device Research Conference Digest*, 1999, pp. 46–47.
23. J. Wang and B. W. Williams, Evaluation of High-Voltage 4H-SiC Switching Devices, *IEEE Trans. Electron Devices*, Vol. 46, No. 3, March 1999, pp. 589–597.
24. J. B. Casady, A. K. Agarwal, S. Seshadri, R. R. Siergiej, L. B. Rowland, M. F. Macmillan, D. C. Sheridan, P. A. Sanger, and C. D. Brandt, 4H-SiC Power Devices for Use in Power Electronic Motor Control, *Solid State Electron.*, Vol. 42, No. 12, December 1998, pp. 2165–2176.

APPENDIX 10.1: ELECTRON CURRENT AT THE COLLECTOR JUNCTION

From Appendix 5.2, Eq. (A5.2.16), the electron current is given by

$$I_{\mathrm{n}W_{\mathrm{b}}} = \frac{b}{1+b} I_{\mathrm{t}} + qAD_{\mathrm{a}} \frac{dn(x)}{dx} \tag{A10.1.1}$$

Representing the carrier concentration $n(x)$ by a linear distribution

$$n(x) = -\frac{p_0 - p_w}{W_b}x + p_w \tag{A10.1.2}$$

we obtain by direct differentiation

$$\frac{dn(x)}{dx} = -\frac{p_0 - p_w}{W_b} \tag{A10.1.3}$$

Substituting for $dn(x)/dx$ from Eq. (A10.1.3) into Eq. (A10.1.1), we get

$$I_{nW_b} = \frac{b}{1+b}I_t + qAD_a \times -\frac{p_0 - p_w}{W_b} = \frac{b}{1+b}I_t - qAD_a \times \frac{p_0 - p_w}{W_b}$$

$$= \frac{b}{1+b}I_t + \frac{qAD_a(p_w - p_0)}{W_b} \tag{A10.1.4}$$

APPENDIX 10.2: TRANSIENT BASE STORED CHARGE $Q_{base}(t)$

For analytical modeling, the transient carrier distribution in the N^-base is approximated as two linear sections. Figure A10.2.1 shows the approximated transient carrier distribution. p_0 and p_w are the carrier concentrations at the collector and emitter ends, respectively, and L_c is a constant distance in the

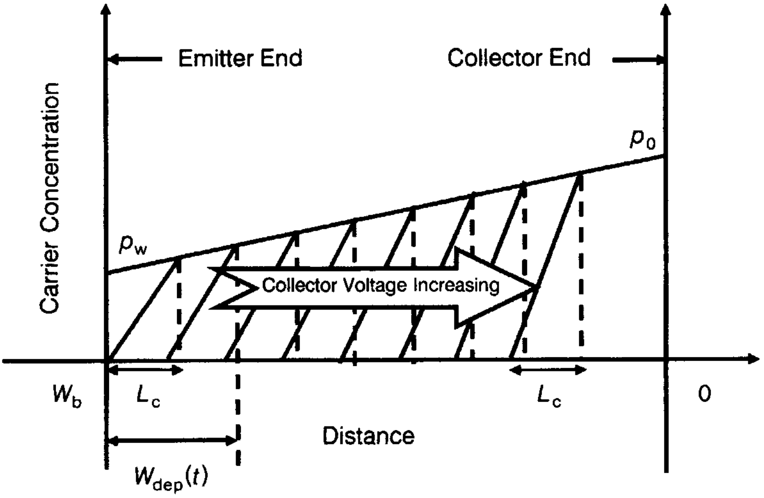

Carrier Concentration
Emitter End
Collector End
p_0
p_w
Collector Voltage Increasing
W_b
L_c
L_c
0
Distance
$W_{dep}(t)$

Figure A10.2.1 Approximated transient carrier distribution.

transient carrier distribution, observed from Fig. A10.2.1. Upon inspection of the diagram, by visualizing the structure of trench-gate IGBT, it follows that transient base stored charge $Q_{\text{base}}(t)$ for a given depletion region depth = Initial base stored charge Q_0 (at the start of the building of the depletion region) – Transient charge Q' due to traversing of carriers across the collector junction (governed by electron extraction or hole sweeping out), that is

$$Q_{\text{base}}(t) = Q_0 - Q' \tag{A10.2.1}$$

Now,

$$\begin{aligned} Q_0 &= \text{Mean charge in the base} - \text{Charge in a constant distance } L_c \\ &= \tfrac{1}{2}qAW_{\text{b}}(p_0 + p_{\text{w}}) - \tfrac{1}{2}qAp_{\text{w}}L_{\text{c}} \end{aligned} \tag{A10.2.2}$$

In the first term the carrier concentration is approximated by $(1/2)(p_0 + p_{\text{w}})$, while in the second term it is $(1/2)(p_{\text{w}} + 0) = (1/2)p_{\text{w}}$. Likewise,

$$\begin{aligned} Q' &= \text{Mean charge in the depletion depth } W_{\text{dep}}(t) \text{ at time } t \\ &\quad - \text{Charge in the length } L_c \\ &= \tfrac{1}{2}qAW_{\text{dep}}(t)(p_{\text{t}0} + p_t) - \tfrac{1}{2}qA(p_{\text{t}0} - p_{\text{t}})L_{\text{c}} \end{aligned} \tag{A10.2.3}$$

where p_{t} = carrier concentration at time t given by

$$p_{\text{t}} = p_{\text{w}} + \frac{p_0 - p_{\text{w}}}{W_{\text{b}}}\{W_{\text{dep}}(t) + L_{\text{c}}\} \tag{A10.2.4}$$

which is the equation of a straight line having gradient $(p_0 - p_{\text{w}})/W_{\text{b}}$ and cutting off an intercept p_{w} from the ordinate. Also,

$$p_{\text{t}0} = p_{\text{t}}|_{W_{\text{dep}}(t)=0} \tag{A10.2.5}$$

APPENDIX 10.3: DEPLETION WIDTH IN THE PRESENCE OF MOBILE CARRIER CONCENTRATION

The sum-total charge in the modulated N^- base region is obtained by adding the mobile hole concentration (p_{m}) to the base doping concentration N_{D}. The mobile hole concentration is expressed as

$$p_{\text{m}} = \frac{I_{\text{t}}}{qAv_{\text{p sat}}} \tag{A10.3.1}$$

where I_{t} is the total device current and v_{psat} is the hole saturation velocity.

Hence, the total base charge is

$$N_D + p_m = \frac{I_t}{qAv_{p\,sat}} \tag{A10.3.2}$$

Then the conventional equation for depletion layer width

$$W_{dep} = \sqrt{\frac{2\varepsilon_0 \varepsilon_s V_a}{N_D}} \tag{A10.3.3}$$

is modified to

$$W_{dep} = \sqrt{\frac{2\varepsilon_0 \varepsilon_s V_a}{N_D + p_m}} \tag{A10.3.4}$$

where $(N_D + p_m)$ is given by Eq. (A10.3.2).

APPENDIX 10.4: MODULATED BASE RESISTANCE (R_b)

The carrier distribution in the N^- base of the IGBT is approximated by a one-dimensional linear distribution in which the carrier concentration decreases from p_0 on the collector side $(x = 0)$ to p_w on the emitter side $(x = W_b$, the base width), as shown in Fig. A10.2.1:

$$n(x) = p(x) = \frac{p_0 - p_w}{0 - W_b}(x - 0) + p_w = -\frac{p_0 - p_w}{W_b}x + p_w \tag{A10.4.1}$$

where $n(x) = p(x)$ for electrical neutrality in the modulated base region. The resultant base resistance is given by

$$R_{base} = \frac{1}{q(\mu_n + \mu_p)A}\int_0^{W_b} \frac{1}{n(x)}\,dx \tag{A10.4.2}$$

where A is the device area. Differentiating Eq. (A10.4.1) with respect to x,

$$\frac{dn(x)}{dx} = -\frac{p_0 - p_w}{W_b} \tag{A10.4.3}$$

Therefore

$$dx = -\frac{W_b}{p_0 - p_w}dn(x) \tag{A10.4.4}$$

Substituting for dx from Eq. (A10.4.4) into Eq. (A10.4.2), we have

$$
\begin{aligned}
R_{\text{base}} &= \frac{1}{q(\mu_n+\mu_p)A} \times -\frac{W_b}{p_0-p_w} \times \int_0^{W_b} \frac{dn(x)}{n(x)} \\
&= -\frac{W_b}{q(\mu_n+\mu_p)A} \times \frac{1}{p_0-p_w} \times \int_{p_0}^{p_w} \frac{dn(x)}{n(x)} \\
&= -\frac{W_b}{q(\mu_n+\mu_p)A} \times \frac{1}{p_0-p_w} \times [\ln n(x)]_{p_0}^{p_w} \\
&= -\frac{W_b}{q(\mu_n+\mu_p)A} \times \frac{1}{p_0-p_w} \times (\ln p_w - \ln p_0) \\
&= \frac{W_b}{q(\mu_n+\mu_p)A} \frac{\ln p_0 - \ln p_w}{p_0-p_w}
\end{aligned}
\tag{A10.4.5}
$$

APPENDIX 10.5: ON-STATE VOLTAGE DROP DUE TO RECOMBINATION IN THE END REGIONS OF THE PIN DIODE IN THE IGBT

The potential drop across the IGBT in the ON state = Ohmic voltage drop across the modulated base resistance (R_b) + MOSFET voltage drop (V_{MOS}) + Voltage drop across the end regions (N^- base/P^+ and N^- base/P base), that is,

$$
V_{\text{ON}} = R_{\text{base}} I_t + V_{\text{MOS}} + \frac{kT}{q} \ln\left(\frac{p_0^2}{n_i^2}\right) \tag{A10.5.1}
$$

representing a series combination of these three components.

To derive the third term of this equation, we note, by Eq. (4.24), that the voltage drop across the N^- base/P^+ collector junction is related to the injected minority carrier density p_0 as

$$
p_0 = p \exp\left(\frac{qV_{P^+}}{kT}\right) \tag{A10.5.2}
$$

where p is the equilibrium minority carrier density in the N^-base. Equation (A10.5.2) is rearranged as

$$
\frac{p_0}{p} = \exp\left(\frac{qV_{P^+}}{kT}\right) \tag{A10.5.3}
$$

giving

$$\ln\left(\frac{p_0}{p}\right) = \frac{qV_{P^+}}{kT} \tag{A10.5.4}$$

or,

$$V_{P^+} = \frac{kT}{q}\ln\left(\frac{p_0}{p}\right) \tag{A10.5.5}$$

Similarly, at the N$^-$ base/N$^+$ junction we have

$$p_w = N_D \exp\left(\frac{qV_{N^+}}{kT}\right) \tag{A10.5.6}$$

where N_D is the N$^-$ base doping concentration. Hence,

$$V_{N^+} = \frac{kT}{q}\ln\left(\frac{p_w}{N_D}\right) \tag{A10.5.7}$$

Adding together Eqs. (A10.5.5) and (A10.5.7), we get

$$V_{P^+} + V_{N^+} = \frac{kT}{q}\left\{\ln\left(\frac{p_0}{p}\right) + \ln\left(\frac{p_w}{N_D}\right)\right\} = \frac{kT}{q}\ln\left(\frac{p_0 p_w}{pN_D}\right) \tag{A10.5.8}$$

The product pN_D is replaced by n_i^2, where n_i is the intrinsic carrier concentration. Then Eq. (A10.5.8) is written as

$$V_{P^+} + V_{N^+} = \frac{kT}{q}\ln\left(\frac{p_0 p_w}{n_i^2}\right) \tag{A10.5.9}$$

For $p_0 = p_w$, this equation becomes

$$V_{P^+} + V_{N^+} = \frac{kT}{q}\ln\left(\frac{p_0^2}{n_i^2}\right) \tag{A10.5.10}$$

APPENDIX 10.6: ENERGY LOSS

This consists of two components:

(i) Energy loss occurs when the collector voltage rises to the switching rail voltage. The energy loss taking place during this period is obtained by integrating the collector voltage $V_C(t)$, giving

$$E_{\mathrm{vr}} = I_t \int_0^{t_{\mathrm{vr}}} V_C(t)\,dt \tag{A10.6.1}$$

where t_{vr} is the time taken by the collector voltage to reach the rail voltage.

(ii) Energy loss also takes place during the current tail phase of IGBT turn-off. The tail energy loss is expressed as

$$E_{\mathrm{tail}} \cong V_{\mathrm{DC}} \cdot Q_{\mathrm{base}}(t)\big|_{V_C(t)=V_{\mathrm{DC}}} \tag{A10.6.2}$$

where V_{DC} is the rail voltage during switching.

The total energy loss is obtained by adding Eqs. (A10.6.1) and (A10.6.2).

APPENDIX 10.7: EXCESS CARRIER CONCENTRATION p_w IN THE N$^-$ BASE OF TIGBT AT THE EMITTER END

To determine the excess carrier concentration (p_w) in the N$^-$ base of TIGBT at the emitter end, the ambipolar diffusion equation is derived and solved with the associated boundary conditions, to obtain the electron and hole current density distributions. Then p_w is expressed in terms of recombination current density J_{rec} and $J_{\mathrm{rec}}(w)$ is found from the hole current density distribution to yield the equation for p_w. For high-level injection, the electron current density in the N$^-$ base is

$$J_n(y) = \frac{b}{1+b}J + qD\frac{dp}{dy} \tag{A10.7.1}$$

and the hole current density is

$$J_p(y) = \frac{1}{1+b}J - qD\frac{dp}{dy} \tag{A10.7.2}$$

In these equations, $p(y) \approx q(y)$ is the excess carrier distribution in the N$^-$ base, J is the total current density, b is the electron-to-hole mobility ratio, and D is the ambipolar diffusion constant. Applying the continuity equation

$$\frac{dJ_p}{dy} = -\frac{qp}{\tau} \tag{A10.7.3}$$

and the quasi-neutrality condition

$$n(y) \cong p(y), \qquad \frac{dn(y)}{dy} \cong \frac{dp(y)}{dy} \tag{A10.7.4}$$

Eqs. (10.7.2) and (10.7.3) yield

$$\frac{dJ_{\mathrm{p}}}{dy} = -qD\frac{d^2p(y)}{dy^2} = -\frac{qp}{\tau}$$

or

$$\frac{d^2p(y)}{dy^2} = \frac{p(y)}{D\tau} = \frac{p(y)}{L^2} \tag{A10.7.5}$$

where L = ambipolar diffusion length = $\sqrt{D\tau}$, D is the diffusion constant, and τ is the ambipolar lifetime. Equation (A10.7.5) is the *ambipolar diffusion equation* with the boundary conditions $p(0) = p_0$ = excess carrier concentration in the N$^-$ base at the collector end, $p(w) = p_{\mathrm{w}}$ = the same at the emitter end. Equation (10.7.5) is of the standard form

$$a\frac{d^2y}{dx^2} + b\frac{dy}{dx} + cy = 0 \tag{A10.7.6}$$

where a, b, and c are constants. The solution to this equation is

$$y = A\exp(m_1x) + B\exp(m_2x) \tag{A10.7.7}$$

where A and B are constants and m_1 and m_2 are the roots of the quadratic equation

$$am^2 + Bm + c = 0 \tag{A10.7.8}$$

which is known as the *auxiliary equation.* The auxiliary equation for Eq. (A10.6.5) is

$$m^2 - \frac{1}{L^2} = 0 \tag{A10.7.9}$$

Therefore

$$m = \pm\frac{1}{L} \tag{A10.7.10}$$

Hence, the solution to Eq. (A10.7.5) is

$$p(y) = A\exp\left\{\left(\frac{1}{L}\right)y\right\} + B\exp\left\{-\left(\frac{1}{L}\right)y\right\} \tag{A10.7.11}$$

To evaluate the constants A and B, the boundary conditions are applied. At $y = 0$, $p(0) = p_0$. So,

$$p_0 = A + B \tag{A10.7.12}$$

At $y = w$, $p(w) = p_{\mathrm{w}}$, giving

$$p_{\mathrm{w}} = A\exp\left\{\left(\frac{w}{L}\right)\right\} + B\exp\left\{-\left(\frac{w}{L}\right)\right\} \tag{A10.7.13}$$

From Eq. (A10.7.12),

$$B = p_0 - A \tag{A10.7.14}$$

Substituting for B in Eq. (A10.7.13), we get

$$\begin{aligned} p_{\mathrm{w}} &= A\exp\left\{\left(\frac{w}{L}\right)\right\} + (p_0 - A)\exp\left\{-\left(\frac{w}{L}\right)\right\} \\ &= A\exp\left(\frac{w}{L}\right) + p_0\exp\left(-\frac{w}{L}\right) - A\exp\left(-\frac{w}{L}\right) \\ &= A\left\{\exp\left(\frac{w}{L}\right) - \exp\left(-\frac{w}{L}\right)\right\} + p_0\exp\left(-\frac{w}{L}\right) \end{aligned} \tag{10.7.15}$$

On rearrangement, we obtain

$$A = \frac{p_{\mathrm{w}} - p_0\exp\left(-\frac{w}{L}\right)}{\exp\left(\frac{w}{L}\right) - \exp\left(-\frac{w}{L}\right)} \tag{A10.7.16}$$

From Eq. (A10.7.14), we have

$$B = p_0 - A = p_0 - \frac{p_w - p_0 \exp\left(-\frac{w}{L}\right)}{\exp\left(\frac{w}{L}\right) - \exp\left(-\frac{w}{L}\right)}$$

$$= \frac{p_0 \exp\left(\frac{w}{L}\right) - p_0 \exp\left(-\frac{w}{L}\right) - p_w + p_0 \exp\left(-\frac{w}{L}\right)}{\exp\left(\frac{w}{L}\right) - \exp\left(-\frac{w}{L}\right)}$$

$$= \frac{p_0 \exp\left(\frac{w}{L}\right) - p_w}{\exp\left(\frac{w}{L}\right) - \exp\left(-\frac{w}{L}\right)} \tag{A10.7.17}$$

Putting the values of A and B from Eqs. (A10.7.16) and (A10.7.17) into Eq. (A10.7.11), we obtain

$$p(y) = \frac{p_w - p_0 \exp\left(-\frac{w}{L}\right)}{\exp\left(\frac{w}{L}\right) - \exp\left(-\frac{w}{L}\right)} \exp\left(\frac{y}{L}\right)$$

$$+ \frac{p_0 \exp\left(\frac{w}{L}\right) - p_w}{\exp\left(\frac{w}{L}\right) - \exp\left(-\frac{w}{L}\right)} \exp\left(-\frac{y}{L}\right)$$

$$= \frac{p_w \exp\left(\frac{y}{L}\right) - p_0 \exp\left(-\frac{w-y}{L}\right)}{\exp\left(\frac{w}{L}\right) - \exp\left(-\frac{w}{L}\right)}$$

$$+ \frac{p_0 \exp\left(\frac{w-y}{L}\right) - p_w \exp\left(-\frac{y}{L}\right)}{\exp\left(\frac{w}{L}\right) - \exp\left(-\frac{w}{L}\right)}$$

$$= \frac{p_0\left\{\exp\left(\frac{w-y}{L}\right) - \exp\left(-\frac{w-y}{L}\right)\right\}}{\exp\left(\frac{w}{L}\right) - \exp\left(-\frac{w}{L}\right)}$$

$$+\frac{p_w\left\{\exp\left(\frac{y}{L}\right)-\exp\left(-\frac{y}{L}\right)\right\}}{\exp\left(\frac{w}{L}\right)-\exp\left(-\frac{w}{L}\right)}$$

$$=p_0\frac{\sinh\left(\frac{w-y}{L}\right)}{\sinh\left(\frac{w}{L}\right)}+p_w\frac{\sinh\left(\frac{y}{L}\right)}{\sinh\left(\frac{w}{L}\right)} \tag{A10.7.18}$$

Substituting for $p(y)$ from Eq. (A10.7.18) into Eq. (A10.7.1) and (A10.7.2), we get

$$J_n(y)=\frac{bJ}{1+b}-\frac{qDp_0}{L}\frac{\cosh\left(\frac{w-y}{L}\right)}{\sinh\left(\frac{w}{L}\right)}+\frac{qDp_w}{L}\frac{\cosh\left(\frac{y}{L}\right)}{\sinh\left(\frac{w}{L}\right)} \tag{A10.7.19}$$

$$J_p(y)=\frac{J}{1+b}+\frac{qDp_0}{L}\frac{\cosh\left(\frac{w-y}{L}\right)}{\sinh\left(\frac{w}{L}\right)}-\frac{qDp_w}{L}\frac{\cosh\left(\frac{y}{L}\right)}{\sinh\left(\frac{w}{L}\right)} \tag{A10.7.20}$$

Carrier concentrations p_0 and p_w are expressed in terms of the recombination current densities in the end regions $J_{rec}(0)$ and $J_{rec}(w)$ as

$$p_0^2=\frac{J_{rec}(0)}{qh_0}=\frac{J_n(0)}{qh_0} \tag{A10.7.21}$$

$$p_w^2=\frac{J_{rec}(w)}{qh_w} \tag{A10.7.22}$$

where for abrupt junctions we obtain

$$h_0=\frac{J_{ns}}{qn_i^2},\qquad h_w=\frac{J_{ps}}{qn_i^2} \tag{A10.7.23}$$

J_{ns}, J_{ps} are saturation current densities. $J_{rec}(0)$ is the electron current density flowing into the P$^+$ region across the collector junction, and $J_{rec}(w)$ is the recombination current density flowing into the accumulation layer across the N$^+$ accumulation layer/N$^-$-base junction.

The hole current density $J_p(w)$ at the emitter end of N$^-$ base consists of two components, one of which $J_{rec}(w)$ recombines in the accumulation layer while the other, $\alpha_{PNP}J$, flows through the P well to the collector short contact; hence

$$J_{\mathrm{p}}(w) = J_{\mathrm{rec}}(w) + \alpha_{\mathrm{PNP}} J \tag{A10.7.24}$$

or

$$J_{\mathrm{rec}}(w) = J_{\mathrm{p}}(w) - \alpha_{\mathrm{PNP}} J \tag{A10.7.25}$$

Substituting for $J_{\mathrm{p}}(w)$ from Eq. (A10.7.20) into Eq. (A10.7.25), we get

$$\begin{aligned} J_{\mathrm{rec}}(w) &= \frac{J}{1+b} + \frac{qDp_0}{L}\frac{1}{\sinh\left(\frac{w}{L}\right)} - \frac{qDp_{\mathrm{w}}}{L}\coth\left(\frac{w}{L}\right) - \alpha_{\mathrm{PNP}} J \\ &= \left(\frac{1}{1+b} - \alpha_{\mathrm{PNP}}\right)J + \frac{qD}{L}\frac{p_0}{\sinh\left(\frac{w}{L}\right)} - \frac{qD}{L}p_{\mathrm{w}}\coth\left(\frac{w}{L}\right) \end{aligned} \tag{A10.7.26}$$

p_{w} is obtained by substituting for $J_{\mathrm{rec}}(w)$ from Eq. (A10.7.26) into Eq. (A10.7.22):

$$\begin{aligned} p_{\mathrm{w}}^2 = {} & \frac{1}{qh_{\mathrm{w}}}\left(\frac{1}{1+b} - \alpha_{\mathrm{PNP}}\right)J + \frac{1}{qh_{\mathrm{w}}} \times \frac{qD}{L}\frac{p_0}{\sinh\left(\frac{w}{L}\right)} \\ & - \frac{1}{qh_{\mathrm{w}}} \times \frac{qD}{L}p_{\mathrm{w}}\coth\left(\frac{w}{L}\right) \end{aligned} \tag{A10.7.27}$$

or

$$p_{\mathrm{w}}^2 + \frac{D}{h_{\mathrm{w}}L}\coth\left(\frac{w}{L}\right)p_{\mathrm{w}} - \frac{1}{qh_{\mathrm{w}}}\left(\frac{1}{1+b} - \alpha_{\mathrm{PNP}}\right)J - \frac{D}{h_{\mathrm{w}}L}\frac{p_0}{\sinh\left(\frac{w}{L}\right)} = 0 \tag{10.7.28}$$

which is a quadratic equation in p_{w} having the solution

$$p_{\mathrm{w}} = \frac{-\frac{D}{h_{\mathrm{w}}L}\coth\left(\frac{w}{L}\right) \pm \sqrt{\begin{aligned} & \frac{D^2}{h_{\mathrm{w}}^2L^2}\coth^2\left(\frac{w}{L}\right) + \frac{4\times1\times1}{qh_{\mathrm{w}}}\left(\frac{1}{1+b} - \alpha_{\mathrm{PNP}}\right)J \\ & + \frac{4D}{h_{\mathrm{w}}L}\frac{p_0}{\sinh\left(\frac{w}{L}\right)} \end{aligned}}}{2\times1}$$

$$= \frac{-\frac{D}{h_w L}\coth\left(\frac{w}{L}\right) \pm \sqrt{\frac{D^2}{h_w{}^2 L^2}\coth^2\left(\frac{w}{L}\right) + \frac{D^2}{h_w{}^2 L^2} \times \frac{4h_w L}{qD^2}\left(\frac{1}{1+b} - \alpha_{\text{PNP}}\right)J + \frac{D^2}{h_w{}^2 L^2} \times \frac{4h_w L}{D}\frac{p_0}{\sinh\left(\frac{w}{L}\right)}}}{2 \times 1}$$

$$= \frac{D}{2h_w L}\left[\frac{-\coth\left(\frac{w}{L}\right) \pm \sqrt{\coth^2\left(\frac{w}{L}\right) + \frac{D^2}{h_w{}^2 L^2} \times \frac{4h_w D\tau}{qD^2}\left(\frac{1}{1+b} - \alpha_{\text{PNP}}\right)J + \frac{4h_w L p_0}{D\sinh\left(\frac{w}{L}\right)}}}{2}\right] \quad \text{(A10.7.29)}$$

APPENDIX 10.8: ON-STATE VOLTAGE DROP ACROSS THE N$^-$ BASE OF IGBT

Since electric field $E(y)$ across the N$^-$ base is related to the potential drop V_{N^-} as

$$E(y) = -\frac{\partial V_{\text{N}^+}}{\partial y} \quad \text{(A10.8.1)}$$

the potential drop V_{N^-} is

$$V_{\text{N}^-} = -\int_0^w E(y)\,dy \quad \text{(A10.8.2)}$$

To substitute for $E(y)$, we recall Eqs. (4.47) and (4.48):

$$J_n = q\mu_n\left(nE + \frac{kT}{q}\frac{dn}{dy}\right) \quad \text{(A10.8.3)}$$

and

$$J_p = q\mu_p\left(pE - \frac{kT}{q}\frac{dn}{dy}\right) \quad \text{(A10.8.4)}$$

Adding together Eqs. (A10.8.3) and (A10.8.4), we get

$$J_n + J_p = q\mu_p\left(pE - \frac{kT}{q}\frac{dn}{dy}\right) + q\mu_n\left(nE + \frac{kT}{q}\frac{dn}{dy}\right) \quad \text{(A10.8.5)}$$

from which, noting that $J_n + J_p = J$ and $n = p$,

$$E = \frac{J}{q(\mu_n + \mu_p)p} - \frac{kT}{q}\frac{\mu_n - \mu_p}{\mu_n + \mu_p}\frac{1}{p}\frac{dp}{dy} \qquad \text{(A10.8.6)}$$

Applying Einstein's equation [Appendix 4.2, Eq. (A4.2.9)], we obtain

$$\frac{D}{\mu} = \frac{kT}{q} \quad \text{or} \quad \mu = \left(\frac{q}{kT}\right)D \qquad \text{(A10.8.7)}$$

Replacing μ_n by $\{q/(kT)\}D_n$ and μ_p by $\{q/(kT)\}D_p$, and p by $p(y)$, we have

$$E = \frac{J}{q(\mu_n + \mu_p)p(y)} - \frac{kT}{q}\frac{\frac{q}{kT}(D_n - D_p)}{\frac{q}{kT}(D_n + D_p)}\frac{1}{p(y)}\frac{dp(y)}{dy} \qquad \text{(A10.8.8)}$$

Substituting for E from Eq. (A10.8.8) into Eq. (A10.8.2), we have

$$V_{N^-} = \int_0^w \left[\left\{\frac{J}{q(\mu_n + \mu_p)p(y)}\right\}dy - \left\{\left(\frac{D_n - D_p}{D_n + D_p}\right)\left(\frac{kT}{q}\right)\frac{dp(y)}{p(y)}\right\}\right] \qquad \text{(A10.8.9)}$$

Let us consider the first term of this equation. Substituting for $p(y)$ from Eq. (A10.7.18) into the first term of Eq. (A10.8.9), we have

$$\text{First term} = \frac{J}{q(\mu_n + \mu_p)}\int_0^w \frac{dy}{p(y)}$$

$$= \frac{J}{q(\mu_n + \mu_p)}\int_0^w \frac{dy}{\dfrac{p_0 \sinh\left(\frac{w-y}{L}\right)}{\sinh\left(\frac{w}{L}\right)} + \dfrac{p_w \sinh\left(\frac{w}{L}\right)}{\sinh\left(\frac{w}{L}\right)}}$$

$$= \frac{J}{q(\mu_n + \mu_p)}\int_0^w \frac{1}{\sinh\left(\frac{w}{L}\right)} \frac{dy}{\left[p_0 \sinh\left(\frac{w-y}{L}\right) + p_w \sinh\left(\frac{w}{L}\right)\right]}$$

$$= \frac{J \sinh\left(\frac{w}{L}\right)}{q(\mu_n + \mu_p)}\int_0^w \frac{dy}{\left[p_0 \sinh\left(\frac{w-y}{L}\right) + p_w \sinh\left(\frac{w}{L}\right)\right]} \qquad \text{(A10.8.10)}$$

Now directing our attention on the second term of Eq. (A10.8.9), we have

$$\text{Second term} = \left(\frac{D_\mathrm{n}-D_\mathrm{p}}{D_\mathrm{n}+D_\mathrm{p}}\right)\left(\frac{kT}{q}\right)\int_0^w \frac{dp(y)}{p(y)} = \left(\frac{D_\mathrm{n}-D_\mathrm{p}}{D_\mathrm{n}+D_\mathrm{p}}\right)\left(\frac{kT}{q}\right)$$

$$\times[\ln p(y)]_{p=p_0}^{p=p_\mathrm{w}} = \left(\frac{D_\mathrm{n}-D_\mathrm{p}}{D_\mathrm{n}+D_\mathrm{p}}\right)\left(\frac{kT}{q}\right)[\ln p_\mathrm{w} - \ln p_0]$$

$$= \left(\frac{D_\mathrm{n}-D_\mathrm{p}}{D_\mathrm{n}+D_\mathrm{p}}\right)\left(\frac{kT}{q}\right)\ln\left(\frac{p_\mathrm{w}}{p_0}\right) \qquad \text{(A10.8.11)}$$

The N$^-$-base voltage drop is obtained by adding together Eqs. (A10.8.10) and (A10.8.11).

The Insulated Gate Bipolar Transistor IGBT Theory and Desig, by Vinod Kumar Khanna
ISBN: 0-471-23845-7 © 2003 by the Institute of Electrical and Electronics Engineers

INDEX

2-D model, carrier distribution, 238-241
3C-SiC, 350
4H-SiC, 1, 350, 526
6H-SiC, 350, 525

Abrupt junction, 157, 165, 353, 359, 405-406
  approximation, 353
  physical model, 158, 229
Absolute maximum ratings of IGBT, 91, 362
Absolute temperature, 161, 205, 291, 432
Acceptor, 157
  doping concentration, 160
  impurity, segregation of, 430
Accumulation, 109
Accumulation layer, 50, 124, 126, 132, 150, 238, 247–248, 502–504, 507, 519
Accumulation layer mobility, 148
Accumulation layer resistance, 123, 125, 127, 148
AC–DC conversion, 570-575
AC motor, 599
AC motor drive, 11, 598
Active region, 51, 78
Adhesive, thermal, 473
Air cooling, 489
Al gate, 117–118
Allotropes, 350
Allotropic modifications, 350
Alpha of a transistor, 66
Alumina, 490
Aluminum metallization, 14, 419–420, 440, 490, 493
Aluminum nitride, 470, 472, 490
Aluminum oxide, 470, 490
Aluminum wire, 494
Ambipolar diffusion coefficient/constant, 160, 174, 177, 211, 246–247, 287–289, 312, 536–537
Ambipolar diffusion equation, 252–253, 258, 266, 289, 537
Ambipolar diffusion length, 71, 160, 174, 234, 237, 247, 249–250, 289, 312, 317, 382, 504
Ambipolar diffusivity, 252, 261, 500, 504
Ambipolar lifetime, 211, 220, 247, 264, 312, 317, 382, 537
Ambipolar mobility, 177
Ambipolar mobility ratio, 252, 288
Ambipolar transport equations, 18, 251, 254, 266–267
Amplification, 132
Amplifying gate, thyristor, 193
Analog circuit, representation of IGBT model equations by, 591
Analytical calculation, 362–364
Analytical simulation, 229
Anisotropic etching, 422, 438
Annealing, 351, 414, 434, 441, 444, 446, 448–449
Anode, 23, 155, 173, 188–189, 192, 358
  current, 160, 190
  junction, 217, 219
Antiparallel diode, 76
APCVD. *See* Atmospheric pressure CVD
Applications of IGBT, 90, 545–608
Arc welding, 603
Armature, 557, 598
Aspect ratio, 244, 364, 368, 502
Asymmetrical IGBT, 38, 40, 43
  latching of, with resistive load, 314–315
Atmospheric pressure CVD (APCVD), 436–437
Atomic lattice layout (A-L-L), 138, 333–337
Auger
  recombination, 176–177, 181
  lifetime, 177
Automotive electronics, 11
Automotive ignition, 89, 600
Auxiliary diode, 28
Auxiliary emitter, 511–512
Auxiliary equation, 537

Avalanche breakdown, 79, 82, 165, 177, 181, 187, 189, 355, 357, 367
  voltage (V_B), 404–406
Avalanche multiplication, 136
Avalanche multiplication coefficient, 352

Back EMF, 551
BAMBI, 364
Band bending, 108, 116
Band flattening, 117
Bandgap, 108, 113, 141, 350
Bandgap narrowing, 181, 366
Bandgap variation with temperature, 366
Band-to-band excitation, 165
Barrier height, 107–108, 112, 114
Barrier metal, 440
Barrier potential, 162, 167
Base, 71
  charge, 533
  contact, virtual, 241–242
  current, 160, 181, 190, 154
  doping profile, 368-372
  widening, 182
  width modulation, 182
Base drive, 104, 184
Base resistance (R_b), 94, 136, 256, 278, 501, 533-534, 591
Base stored charge, transient, Q_{base} (t), 531–532
Base transit time, 71
Base transport factor, 161, 181, 227, 249, 317, 352, 373, 403
Basis function, 299
  finite element, 297
Battery, 552, 556, 558
BCD (Bipolar, CMOS, DMOS), 478
Behavior modeling, 229
Beryllium oxide, 470, 490
Beta, 71, 589
Beveling, 48
Bidirectional blocking mode, 524
Bidirectional conduction, 522
Bimetallic effect, 487–488
Bipolar components of IGBT, 155
Bipolar junction transistor, 2, 9, 104, 106, 180, 229, 242, 526, 588
Bipolar second breakdown, 134, 136
Bipolar transistor-DMOSFET model of IGBT, 232, 241, 243, 245, 381
  with device-circuit interactions, 251
Bird's beak effect, 446, 505–506
BJT. *See* Bipolar junction transistor
Blocking characteristics, 376, 379
Blocking voltage, 190
Body bias coefficient, 136
Body diode, 104
Body resistance, 130-131
Body terminal, MOSFET, 103
Boltzmann constant, 141, 146, 159–160, 163, 205, 291, 432
Boost converter, 553–556
Boron implantation, 40
Braking mode, 557
Breakdown field, 350
Breakdown models, 357–361
Breakdown voltage, 9, 48, 82, 92, 127, 132, 134, 161, 164–166, 187, 189–190, 232, 352-357, 359, 367, 386, 389–392, 395, 421, 518
Bridge circuit,
  Half-bridge, 22, 594, 596
  Full-bridge, 22, 595–596
  Three-phase bridge, 22, 479, 491–492, 568
Bromine, 438
Buck-boost converter, 556–557
Buck converter, 546–553
Buffer layer, 16, 40, 66, 318, 381–382, 422, 428, 441, 486, 502, 516, 523
  structure optimization, 381–385
Built-in potential, 83, 136, 157, 161–163, 252, 307, 319, 510, 520, 523
  disappearance of, 280
Bulk mobility, 366–367
Bulk semiconductor potential, 146, 152
Bulk potential, 110
  semiconductor, 146, 152
Buried layer, 478, 511
Buried oxide, 516
BV_{CEO}. *See* Collector-emitter breakdown voltage with base open
BV_{CER}. *See* Collector-emitter breakdown voltage with resistance R connected between base and emitter
BV_{CES}. *See* Collector-emitter breakdown voltage with base shorted to emitter
BV_{CYL}/BV_{PP} equation, 406–410

CAD (Computer aided design), 443
CAD simulation hierarchy, 362–364
Capacitance, 86, 92, 112, 130–132, 160, 162–163, 172, 193, 256, 258, 263, 362, 380, 486, 589, 591
Capacitance-voltage characteristics, 113
Capacitive displacement current, 193
Carrier concentration, 173
Carrier density, 174, 178

Carrier diffusion, 201–202, 206
Carrier distribution, 173, 175–176, 232–235, 238–241, 500
Carrier drift, 201–202, 206
Carrier lifetime, 39, 67, 71, 72, 74, 76, 86, 134, 156, 161, 174, 177, 229, 232, 246, 350, 352, 356, 398, 500
Carrier profile, 239, 264
Cascode configuration, 5
Casing, 156, 175
Catenary carrier distribution, 233
Cathode, 23, 155, 173, 188–189, 192–193, 358
  current, 160
  junction, 217, 219
Caughey-Thomas equation, 367
Cell, 12
  geometry effect on latchup, 328
  pitch, 125, 129
  size, 128
  width, 125, 150, 239, 338
  window, 138
Cellular layouts, 138
Ceramic insulator, 156
Channel, 101, 122, 148, 236, 248, 306, 317
  charge, 122
  charge distribution, 120
  conductance/conductivity, 111, 120, 130
  counterdoping, 325
  density, 87, 106, 125, 127, 324
  length, 106, 122, 129, 198, 231, 246, 256, 364, 368–376, 411, 479
  length reduction, 250
  mobility, 325
  opening, 198
  pinch-off, 52, 120, 122
  resistance, 121–123, 126–127, 129–130, 147–148, 326, 372, 398
  width, 119, 198, 246, 256, 334, 364, 378
Channeling, ion implantation, 412
Channel length effect on latching current density, 372–376
Channel stopper, 394
Characteristic equation, 208
Characteristic impedance, 579
Charge-control model, 71
Charge-control phenomena, 16
Charge-control relation, 182
Charge distribution, 116–117
Charge-neutrality condition, 110, 162, 173, 175–176, 287
Chemical vapor deposition (CVD), 13, 436–437
  atmospheric pressure CVD (APCVD), 436–437
  low-pressure CVD (LPCVD), 436–437
  plasma-enhanced CVD (PECVD), 420, 436–437
Chip mounting, 475
Chlorine, 437–438
Chopper, 63, 68, 266, 491–492, 545, 556
Circuit diagram symbol
  IGBT, 23
  P-N diode, 156
  thyristor, 188
Circuit modeling, 17, 230, 589
Circuit simulation, 17, 241, 587
Circular array, 138
Circular cell, 138, 331–332
Clamp diodes, 89, 90
Clamped inductive load, 266
Clamped inductive load current (I_{CELM}), 91
Clamped inductive load (CIL) switching, 81, 133, 500
Classification of faults in a converter, 480
Closed-loop control, 514, 592, 598–600
Cluster cell, 520
Clustered IGBT in homogeneous base technology (HB-CIGBT), 520–522
CMOS, 476, 478
Coefficient of thermal expansion, 488, 490–491
Collector, 23
  current, 160, 180, 182, 190, 221–222, 225, 254
Collector-base depletion capacitance, 258
Collector-base space-charge concentration, 256
Collector-emitter breakdown voltage with base open (BV_{CEO}), 187, 189, 403
Collector-emitter breakdown voltage with base shorted to emitter, 187, 189
Collector-emitter breakdown voltage with resistance R connected between base and emitter, 187, 189
Collector-emitter redistribution capacitance, 256
Collector injection efficiency, 502
Collector multiplication factor, 160, 181
Collector output resistance, 249–251
Collector-shorted LIGBT, 515
COMFET (see Conductivity-modulated FET)
Common-base current gain, 66, 161, 181–183, 190, 192, 225
Common-emitter current gain (h_{FE}), 71, 160–161, 181–182, 185, 241
Commutation circuit, 584
Complementary devices, 45
Complementary error function (erfc), 431–432

Complementary half-bridge inverter circuit, 46
Complex *t*-plane, 239
Concentration gradient, 175, 201
Conduction band, 108, 141, 441
Conduction angle, 564-565
Conduction losses, 11, 40, 482
Conductivity, 202
Conductivity-modulated base resistance, 256, 278
Conductivity-modulated FET, 5
Conductivity modulation, 50, 87, 173, 176, 182, 239, 376, 505, 518, 520, 523
Conformal mapping, 239
Contact potential, 112, 114–115, 117
Contact resistance, 505
Continuity equation, 169, 174–175, 177, 180, 206–207, 209, 232, 286, 357, 365, 536
Control section, 592
Conventional current, 203
Conversion. *See* DC-to-DC conversion, DC-AC inversion, AC-DC conversion
Converter, 104, 189, 577, 592
  classification of faults, 480
  DC-to-DC, 545–557
Cooling, 489
  air cooling, 489
  evaporative cooling, 489
  liquid cooling, 489
Cool-MOSFET, 4
Co-ordinate system, 242, 251
Co-pak, 8
Copper, 440, 490
Copper deposition, 470-472
Copper-on-ceramic (COC), 470-472
Copper thick film (CTF) technology, 470–472
Correction function, 588
Counterdoping of channel by ion implantation, 325
Counter EMF, 551
Coupled effect, 502
Critical design parameters of field rings, 385–388
Critical processing steps, VDMOSFET IGBT fabrication, 421
Cross-over time, 57, 63, 68, 70
Current crowding, 81, 398
Current density, 128, 160, 169, 173, 176, 178, 180, 201–202, 204, 207, 219, 230–232, 237–238, 287, 536, 540
  equation, 217
Current-boundary phase, 81, 263-264
Current composition ratio, 514
Current-controlled voltage source (CCVS), 18, 588
Current gain, 160, 180, 193-195, 247, 249–250, 306–307, 310, 312, 317, 352, 364, 373, 376, 382, 588
Current gate drive circuit of IGBT, 24, 25
Current generator, 130–131
Current induced latch up, 78
Current plasma, 193
Current rating, 155
Current sense emitter, 479
Current sensor, 514–515
Current sharing, 467
Current source, 229, 589
  Current source inverter, 599
Current-voltage characteristics, 9, 50, 83, 119, 167, 198, 230–231, 373
Current-voltage equation, 167
Cut-off condition, 184
Cut-off region, 51, 78, 187–188
Cylindrical junction, 335, 386, 406–410
Czochralski (CZ) silicon, 349, 412

Dangling bonds, 113, 351
Darlington transistor, 2, 5, 241, 601
DB-IGBT. *See* Dynamic N-buffer insulated gate bipolar transistor
DC-AC inversion, 558–570
DC chopper, 545, 598
DC-DC conversion, 545–557
DC-DC converter, soft switching, 576–582
DC loop inductance, 480
DC motor, 598
DC motor drive, 598
DC series motor, 598
Dead time, 560
Deal-Grove model of thermal oxidation, 447, 453–457
Debye length, 110, 143
  extrinsic, 143
Deep P^+ diffused region, 136, 318–322, 324, 342
Defined variables, 590
Degenerately doped poly Si gate, 117
Delay angle, 93, 570
Delay time, 56, 63, 68, 91, 92, 132, 161, 195–196, 589
Depletion, 109
  approximation, 357
  charge, 164
Depletion layer, 109, 112, 198, 226–227, 264, 395, 406
  capacitance, 160, 162–163
  curvature, 393
  spreading in IGBT, 44, 49
  thickness, 149

  width, 125, 161–163, 165, 170–171, 198, 250, 252, 314, 406, 408, 500, 532–533
Depletion mode MOSFET, 101
Depletion region, 159, 162–164, 170, 197, 207, 221–224, 359, 385, 507
  emitter, 182
Depletion region logic, 358
Deposition, 444, 447
  epitaxial, 428–429
Design approaches of field ring, 388–390
Design concepts, novel, 499
Design considerations, IGBT unit cell, 349–410
Design, IGBT,
  analytical calculations and numerical simulations, 362–364
  reduction of MOS capacitances, 377–381
Design of field ring 392–393
Design of IGBT cell, Choice of, 328–338
Design parameters (critical) of field rings, 385–388
Design software, 364–365
Design of vertical structure, IGBT, 349–361
Dessis command file, 365
DESSIS-ISE, physical models, 365–368
Developer, 435
Device-circuit interactions, 229, 251
DG-IEGT. *See* Double-gate injection-enhanced gate transistor
di/dt, 8, 171, 193, 480, 482
Dichlorosilane, 428, 437
Die attach, 494
Die bonding, 442
Dielectric isolation, 475–477, 516
Dielectrically isolated LIGBT (DI-LIGBT), 515–516
Diffusion, 201, 204, 229, 458
Diffusion approximation, 288
Diffusion coefficient, 159, 168, 170, 183, 211, 224–225, 247, 287, 291, 431–432, 458
Diffusion constant, 160, 177, 204
Diffusion current, 159, 170, 182, 201, 204
Diffusion equation, 168, 220
Diffusion of free carriers, 157
Diffusion length, 48, 160, 169, 171, 176, 178, 181, 183, 207, 220, 224, 226, 291, 353, 406
Diffusion-limited region, silicon growth rate, epitaxy, 428
Diffusion potential, 254
Diffusion profile, 458
Diffusion, thermal, 13, 444, 447, 458
Diffusivity, 177, 204, 432, 454, 505
DI-LIGBT. *See* Dielectrically isolated LIGBT
Diode. *See* P-N junction
  knee/kink, 231
DIOS-ISE, 365
Direct bond copper (DBC), 493
Direct copper bonding (DCB), 470–472
Direct writing on wafer, 436
Discrete pulse modulation, 585
Discrete transistor, 180
Disk-type package, 157
Displacement capacitance, 263
Displacement current, capacitive, thyristor, 193
Diverter region, P^+, 509–510
  in IGBT cell structure, 338–340
DMOSFET. *See* Double-diffused MOSFET
DMOS-IGBT, 86
Donor, 157
  doping concentration, 160
Donor impurity, segregation of, 430
Doping profiles, 445
  emitter and base, 368–372
Double diffused transistor, 179–180
Double-diffused MOSFET, 3–5, 104–106, 123, 125, 127–129, 136, 138, 230, 242, 248, 340, 480, 482, 527
Double-gate injection-enhanced gate transistor (DG-IEGT), 523–525
Double trench-gate IGBT, 505
Drain, 101, 197–198
Drain resistance, 119
Drain-source capacitance, 132
Drain-source current, 160
Drain-source overlap depletion capacitance, 257
Drain-source voltage drop, 255
Drift, 204, 229
Drift current, 119, 159, 201, 204
Drift-diffusion model, DESSIS, 365
Drift region, 8, 124–126, 129, 149, 231–233, 236–238, 264
  resistance, 126, 129, 149–150
Drift velocity, 202, 205
Drive-in, 419, 444, 431–432, 461–463
Driving mode, 556
Dry oxidation, silicon, 429–431
Duty cycle/ratio, 20, 84, 94, 551–552, 556
dV/dt, 3, 27, 63, 73, 76, 81, 93, 193, 263, 265–266, 421, 479–480, 582, 587
Dynamic electrothermal model of IGBT, 270–280
Dynamic latching, 303, 313, 308–317, 342, 512
Dynamic model, IGBT, 256, 257
  SPICE model, 589
Dynamic N-buffer insulated gate bipolar transistor (DB-IGBT), 505–509
Dynamic power loss, 133

Dynamic properties, 240
Dynamic to static latching current ratio, 311–312, 314, 316

Early tail phase, inductive load turn-off, 269
Ebbers-Moll model of a transistor, 182
Edge termination, 89, 156, 484
Effective density of states, 141, 146
Effective low-doped base doping concentration, 255
Effective N^+ emitter length, 517–518
Effective ON-time, 551
Effective segregation coefficient, 430
Einstein's equation, 159, 171, 204, 543
Electrical and thermal quantities, relationship between, 274
Electrical destruction, 280
Electrical network, 271
Electric field, 159–160, 162–164, 175, 177, 201, 207, 236, 238, 359–360, 395, 433, 520, 542
Electric field crowding, 86, 138, 385, 388, 520
Electric field distribution, 175, 395–396
  IGBT, 43, 236
  MOS structure, 110, 117
Electromagnetic interference (EMI), 62, 483–484, 570, 575, 582, 586
Electromigration, 398, 440
Electron affinity, 107–109, 112
Electron beam lithography, 436
Electron current at the collector junction, 530–531
Electron diffusion, 157
Electronic charge, 161–162
Electronic current, 203
Electronic home, 604
Electron irradiation, 11, 65, 177, 422, 440–441
Electron lifetime, 207
Electrostatic potential, 159
Electrothermal simulation, 229
  Temperature dependency of parameters, 276–277
EMI. *See* Electromagnetic interference
Emitter, 23
  current, 160, 180, 221–222, 225
  length, 333, 375
Emitter ballasting, 106
Emitter-base diffusion capacitance, 256
Emitter-base junction capacitance, 256
Emitter doping profile, 368–372
Emitter injection efficiency, 71, 161, 181, 246, 403, 502
Emitter length, 518
Encapsulant, 493
Encapsulation, 475, 486
Energy balance analysis, 279
Energy-band diagram, 108, 115, 505, 508
Energy band gap, 108, 350
Energy dissipation/loss, 63, 64, 68
Energy gap, 152, 160, 165
Energy loss, 501, 535–536
Enhancement-mode MOSFET, 101, 102
Epitaxial diode, 61
Epitaxial layer, 13, 40, 182, 395–397, 416, 518
Epitaxial transistor, 180, 182
Epitaxy, 180, 412, 428–429, 444, 447, 450
Epoxy, 472, 493
Equations section, 592
Equilibrium segregation coefficient, 430
Equipartition of energy, 204
Equivalent conductive network construction for 1-D linear element, 297–302
Equivalent circuit, 14, 130–131, 230
Error amplifier, 600
Etching, 444, 447
Etching profile, trench, 439
Eutectic bonding, 449
Evaporative cooling, 489
Excess carrier base charge (Q_B), 252–253, 256, 258–260
Excess carrier concentration, 167, 252, 504, 537
  in the N- base of TIGBT at the emitter end, 536–542
Excess carrier distribution, 258, 536
Extinction angle, 570
Extraction of model parameters, 280
Extraction phase, inductive load turn-off, 267
Extrinsic Debye length, 143
Extrinsic semiconductor, 141–142

Fabrication of IGBT, 40, 448–449
Fabrication, trench-gate IGBT, 422–427
Fall time, 57, 62, 64, 70, 91, 92, 132, 187, 589
False turn-on switching, 193
Fast-recovery diode, 60, 155
Fast surface states, 351
Fast thyristor, 189
Fault current, 482
FBSOA. *See* Forward-biased safe operating area
Feedback capacitance, 380
Feedback diode, 560
Fermi-Dirac statistics, 141
Fermi level E_F, 108–110, 116, 120, 141, 146
FET. *See* Field-effect transistor
Fick's law, 443, 458

Field circuit, DC motor, 557, 598
Field-effect transistor, 11, 527
Field oxide, 135, 412, 416, 422, 448
Field plate, 393–394
  design, 385–394
  surface charge effects, 393–394
Field ring, 90, 364, 416, 422
  critical design parameters, 385–388
  design, 385–394
  design approaches, 388–390
  design by making electric field profile uniform, 392–393
  iterative optimization of spacing, 391–392
  numerical simulation of breakdown voltage of PIN diode, 390–391
  spacing optimization, 391–392
Finite element basis function, 297
Finite element method, 274–275
Firing angle, 570
Fixed oxide charge, 112–113, 137
Flat-band
  condition, 108, 115–117
  voltage, 114–115, 117, 134, 440
Flat package, 485
Floating condition, 520
Floating field rings. *See* Field rings
Float-zone (FZ) silicon, 349, 412
Fluence, 67
Fluorine, 437–438
Fluorocarbon, 489–490
Flux, 453–454
Flyback converter, 592–593
Forced h_{FE}, 185
Forward-bias, 167–168, 176, 196
Forward-biased safe operating area (FBSOA), 79, 334, 339, 513, 519
Forward blocking mode of IGBT, 48
Forward blocking state, thyristor, 82, 190–191, 193, 196
Forward breakover, 193
Forward conduction characteristics, 241, 243–245, 372–375, 518
Forward conduction mode of IGBT, 48, 524
Forward characteristic, 377
Forward conducting state, thyristor, 190–191
Forward voltage drop, 21, 39, 84, 133, 176–177, 236, 243, 245, 247, 255, 339, 368–372, 398, 421, 441, 510, 518
Fourier's law, 278
Four-quadrant switch, 525
Free carrier distribution, 232
Free-wheeling diode, 59, 70, 104, 474–475, 479, 484, 486
Free-wheeling interval, 567
Frequency modulation, 545
Fretting, 488
Friction braking, 557
Full-bridge converter, 22, 593–596
Full-bridge inverter, single phase, 491–492, 561–565

GaAs, 350
Gain-enhanced MOSFET, 5
Gain factor, 121
Galerkin finite-element projection, 275
Gamma-ray irradiation, 86, 134
Gap filler, 473
Gate, 23, 101, 189, 197–198
  length, 125, 198, 239–240
Gate-drain capacitance, 132–133, 258
Gate-drain overlap depletion capacitance, 256
Gate drive voltage supply, 133
Gate driving circuits of IGBT, 23, 479
Gate oxidation, 448
Gate-oxide capacitance, 258
Gate oxide thickness reduction, 326–328, 520
Gate-source capacitance, 132, 258
Gate turn-off thyristor (GTO), 2, 6, 526–527
Gate-voltage oscillations, 80
Gaussian function, 433
Gauss' law, 110, 163
GEMFET. *See* Gain-enhanced MOSFET
Generation current, 160, 170, 207
Generation rate, 206, 287
Generation-recombination centers, 86
Generation-recombination mechanisms, 206
Global matrices, 297
g_{m}. *See* Transconductance
Gradual-channel approximation, 119
Grid, 275, 445
Gold, 420
Gold-silver eutectic, 490
Gold wire, 442
GTO. *See* gate turn-off thyristor
Gummel-Poon model, 182

Half-base width, 176
Half-bridge converter, 22, 594, 596
Half-bridge inverter, single phase, 491–492, 558–561, 603
Half poly Si gate length, 248
Half unit cell, 149–151, 239, 249, 330, 520
Half unit cell width, 248
Half-width of base region, 174, 178
Half-width of intrinsic region, 160
Halogen, 437

Handling of IGBT, 23
Hard switching, 72, 81, 479, 576
HB-CIGBT. *See* Clustered insulated gate bipolar transistor in homogeneous base technology
Heat conduction in a plane wall, 278–279
Heat diffusion equation, 270–272, 278
Heat sink, 156, 489, 493
Heavily doped semiconductor, 177
Helium implantation, 61, 441
Hermetic package, 443
Hexagonal array, 138
Hexagonal cell, 138
h_{FE}. *See* Common-emitter current gain
High-field region, 182
High-level injection, 49, 67, 173, 182, 236, 255, 258, 291, 536
High-level lifetime, 161, 174, 220, 247
High-temperature operation of IGBT, 83
High-voltage IGBT cell, 368–371
Hockey-puck package, 156–157
Holding current, 190
Hole current, 221
Hole diffusion, 157
Hole distribution profile, 234–235
Hole-injecting layer, 12
Hole lifetime, 207
Hot spot, 106, 398
h-parameter, 240
HVDC, 2, 7
Hydrodynamic model, DESSIS, 365
Hydrogen annealing, 113, 351
Hysterisis, 37, 38

IEGT. *See* Injection-enhanced IGBT
IGBT. *See* Insulated gate bipolar transistor
IGBT cell design, Choice of, 328–338
IGBT cell structure with diverter region, 338–340
IGBT driving, 479
IGBT fabrication, process and assembly sequence, 448–449
IGBT-inverter-base welding machine, 603
IGBT models
  bipolar transistor-DMOSFET model, 232, 241, 243, 245
  Parallel and coupled PIN diode-PNP transistor model of carrier distribution in the ON-state of trench IGBT, 502–505
  P-I-N rectifier-DMOSFET model, 230
IGBT power modules
  built-in protection schemes, 481
  di/dt, 482
  features, 486, 493–494
  flat-packaged, 484–486
  heat sinks and cooling, 489–490
  material requirements, 490–491
  parasitic components of package, 482–484
  reliability issues, 486–489
  specification sheet, 495
  state-of-the-art, 491–492
IGBT unit cell, design considerations, 349–410
IGCT. *See* Integrated gate commutated thyristor
Image charge, 116
Image plane, 239–240
Impact generation, 82
Impact ionization, 81, 177, 357, 359
Impulse, 205
Impurity diffusion profiles of IGBT, 41–43
Impurity gradient, 165
Impurity scattering, 367
Induction heating, 604
Inductive load turn-off, 263–264
Injection efficiency, 372, 381–382
Injection-enhanced IGBT (IEGT), 6, 86, 523
Injection enhancement, 88, 519
Injection molding, 443
Input capacitance, 18, 23, 25, 132, 134, 589
Input impedance, 243
Input resistance, 131
INSPECT, 365
Instabilities, 80
Insulated gate bipolar transistor, 1, 5, modules, 6, 7, 465–498
Insulated metal substrate (IMS), 470, 493
Insulating ceramic substrate, 472
Insulation resistance, 488
Integral charge-control relation, 182
Integral cycle control, 570
Integrated gate commutated thyristor (IGCT), 2, 7, 527
Integrated power module, 491
Integrated Systems Engineering-Technology Computer Aided Design (ISE-TCAD), 365
Intelligent power module, 491
Interdigitated gate-cathode geometry, thyristor, 193
Interelectrode capacitances, MOSFET, 131
Interface fast states, 134
Interface material, 471
Interface-state charge, 113, 114, 117, 326
Intrinsic carrier concentration, 83, 84, 137, 141, 146, 152, 160, 162–163, 170, 276, 350, 366, 535
Intrinsic carrier density, 152, 162, 170
Intrinsic Debye length, 110

Intrinsic energy level (E_I), 108–110, 146
Intrinsic Fermi level, 141
Intrinsic or I-region, 172–173
Intrinsic semiconductor, 146
Inversion, 109, 111
Inversion layer, 109, 112, 119, 122, 508
  mobility, 148
Inverter, 558
  soft switching, 582–586
  three-phase full-bridge, 568
Inverter thyristor, 2
Involute pattern, thyristor, 193
Ion-beam lithography, 436
Ion implantation, 13, 40, 353, 394–395, 416, 418, 420, 431, 433–435, 444, 447, 512–513, 516
  for counterdoping of channel, 325
Ion implanted region as high-voltage termination, planar diode, 394–395
Ionization coefficient, 360, 402
Ionization integral, 360, 367, 391, 404
Ionization rate, 81, 404
I-region, 172–173
ISE-TCAD. *See* Integrated Systems Engineering-Technology Computer Aided Design
Isolation techniques, 475–478
  Dielectric isolation, 475–477
  P-N junction isolation, 475–478
  Self-isolation, 475–477
Isothermal simulation, 229
I-V characteristics, 9, 231, 241, 245, 249

JFET. *See* Junction field-effect transistor
JFET effect, 239, 339
Junction capacitance, 160, 172
Junction curvature effect, 86
Junction field-effect transistor, 197–198, 527
  region, 86, 123–126, 128, 149, 155, 234, 236, 247–248
  region resistance, 124, 129, 149, 398
Junction isolation, PN, 475–478
Junction temperature (T_j), 26, 78, 85, 91

Kelvin scale, 159
Kirchhoffs law, 184, 469, 547, 555, 560, 563, 581
Kirk effect, 182

Laplace equation, 239, 358
Laplace transform, 459
Laser beam writing, 436
Latch, 303
Latching, 14, 78, 81, 88, 372
Latching current density, 320, 324, 326–327, 337, 341, 382, 384–385, 512–513, 518
  atomic lattice layout, 337
  circular cell, 337
  effect of channel length, 372–376
  linear cell, 337
  trench-gate IGBT, 340–341
Latching current ratio, 311–312, 314, 316, 342
Latching of asymmetric IGBT with resistive load, 314–315
Latching of symmetrical IGBT with resistive load, 308–313
Latching of symmetrical IGBT with inductive load, 315–317
Latching prevention measures, 317–340
  counterdoping of channel by ion implantation, 325
  deep P^+ diffused region, 136, 318–322
  diverter region in IGBT cell structure, 338–340
  gate oxide thickness reduction, 326–328
  IGBT cell design, Choice of, 328–338
  minority-carrier by-pass cell design, 324
  shallow P^+ diffusion incorporation in IGBT cell, 322–324
Latch-up of parasitic thyristor in IGBT, 303–348
Lateral current sensing ratio, 515
Lateral diffusion, 148
Lateral DMOSFET, 35, 37, 103, 120
Lateral IGBT in thin silicon-on-insulator (SOI) substrate, 516–517
Lateral IGBT (LIGBT), 35, 36
  dielectrically isolated, 515–516
  high-temperature latch-up immunity, 511–512
  improved structure, 513–514
  RESURF concept, 395–398
  with integrated current sensor, 514–515
  with reverse blocking capability, 509–511
Lateral P-base resistance, 326
Lateral trench-gate bipolar transistor (LTGBT), 517–518
Law of equipartition of energy, 204
Law of the junction, 168
LDMOS. *See* Lateral DMOSFET
Leadframe, 471–472
Lead-tin solder, 490
Leakage current, 92, 135, 170–171, 421, 441, 476
  surface, 182
Lifetime, excess carrier, 39, 67, 71, 72, 74, 76, 86, 134, 156, 161, 174, 177, 183, 207, 220, 225, 229, 263–264, 441

Lifetime, excess carrier (*Continued*)
  ambipolar, 211, 220, 247, 264, 312, 317, 382, 537
  Auger recombination, 177
  electron, 207
  hole, 207
  killing of, 16, 55, 61, 66, 67, 312–313, 317
  killing of, axial, 61
  killing of, localized, 61, 66, 421–422, 441, 502
LIGBT. *See* Lateral IGBT
Linear cell, 123–124, 127, 147–149, 231, 244, 329–330, 338
Linear rate coefficient, 457
Linear profile, 264
Linear region, 122, 197–198, 255
Linear region MOSFET transconductance parameter (K_{plin}), 277
Linearly graded junction, 165
Liquid cooling, 489
Liquid nitrogen, 490
Lithography, 435–436, 478
  electron-beam, 436
  ion-beam, 436
  optical, 435–436
  X-ray, 435–436
Load inductance, 262
Load line, 184–185
Load resistance, 184, 262
Local declarations, SABER template, 590
Logic circuit, 111, 465–466
Longitudinal electric field, 119, 122
Low-level injection, 258
Low pressure CVD (LPCVD), 422, 436–437
LPCVD. *See* Low pressure CVD
LTGBT. *See* Lateral trench-gate bipolar transistor

Mask, 435–436
Mass transfer coefficient, 454
Mass-transport region, silicon growth rate, epitaxy, 428
Matthiessen rule, 367
Maximum electric field (E_{m}), 160, 164
Maxwell-Boltzmann statistics, 141
MCT. *See* MOS-controlled thyristor)
Mean free path, 202–203
MEDICI, 364
Medium voltage IGBT cell, 369–371
Mesa region, 505, 507–508, 519
Mesa width, 126–127
Mesh, 275, 392, 445
  boundary file, 365
  command file, 365
  refinement criteria, 358
  skeleton, 358
  specification files, 358
MESH-ISE, 365
Metallization, 439–440, 475, 493–494
Metallurgical junction, 159
Metal-oxide-semiconductor field-effect transistor, 2–5, 9, 103–107, 136, 229–231, 241–244, 248–249, 256–257, 277, 317, 420, 526–527, 545, 587–589
Microplasmas, 106
Microprocessor, 82
Microwave oven, 604
Miller amplification, 379
Miller capacitance, 92, 131
Miller effect, 132, 379
Miller indices, 351
Minority carrier
  diffusion coefficient, 181
  diffusion length, 181
  lifetime. *See* Lifetime
Minority-carrier by-pass cell design, 324
Minority-carrier concentration, 167–168
Minority-carrier storage, 107
Mobile carrier concentration, 532
Mobile ionic charge, 112, 113, 429
Mobility, 83, 86, 119–120, 136–137, 159, 161, 174, 177, 202, 204–205, 246–247, 276, 350, 366–367, 505, 525
  in accumulation layer, 248
  in channel, 134
  in inversion layer, 148, 246, 248, 350
  on surface, electron, 129, 256
  ratio, 504, 536
Model of carrier distribution, 238
Model parameter extraction for circuit analysis, 280
Modeling of IGBT, 229
Modulated base resistance (R_{b}), 501, 533–534
Module package, 485
Modules, IGBT. *See* IGBT power modules)
Molding, plastic transfer, 475
Molybdenum, 472, 484, 487
Monocrystalline silicon, 114
Monte-Carlo simulation, 447
MOS capacitance reduction in IGBT design, 377–381
MOS-controlled thyristor (MCT), 6, 362, 520
MOSFET. *See* Metal-oxide-semiconductor field-effect transistor
MOS-SCR, 6
MOS second breakdown, 134, 136
MOS structure, 107–108
MOS turn-off thyristor (MTO), 526

MPWM. *See* Multiple pulse width modulation
Multichip module, 484
Multiple pulse width modulation (MPWM), 567
Multiple surface shorts (MSS) cell, 328–329, 334
Multilplication coefficient *M*, 402–403
Multistability, 37, 38

N-buffer layer structure optimization, 381–385
N-channel IGBT, 45
N-channel MOSFET, 102, 112, 119, 127, 134–135
Negative gate capacitance, 80, 89
Negative photoresist, 435
Negative resistance, 190
Net-list, 590
Neutron irradiation, 86, 134
Neutron-transmutation doped (NTD) silicon, 350, 412
Newton method, 365
NMOS transistor. *See* N-channel MOSFET
Nodal network, 275
Nodal point, 275
Node, 275
Non-degenerate semiconductor, 141
Nonisothermal simulation, 229
Nonpunchthrough IGBT, 38, 39, 72, 74, 83, 84, 165, 167, 412, 421, 494
Non-quasi-static charge state equation, 259–260
Non-quasi-static term, 259
Non-quasi-static voltage state equation, 259–260
Non-self-aligned trench IGBT (NSA-TIGBT), 505
Nonstationary transport, 365
Normalized length parameter, 239
Normally off MOSFET, 101
Normally on MOSFET, 101
NPN transistor, 14, 179–180, 190, 192–194, 306, 317–319, 523, 525
NPT-IGBT. *See* Nonpunchthrough IGBT
NSA-TIGBT. *See* Non-self-aligned trench IGBT
N-type material/semiconductor, 157, 201
Nucleation, 428
Numerical simulation, 229, 239, 362–364
  of breakdown voltage of PIN diode, field rings, 390–391

OFF state, 184, 189–191
OFF-state loss, 21, 523
Ohmic contact, 180, 227
Ohmic voltage drop, 175
Ohms law, 469
One-dimensional equivalent thermal network, 299
One-dimensional linear element, 299
On-off control, 570
ON-resistance, 9, 11, 87, 107, 124–125, 127, 138, 239, 364
  model of, 123
ON state, 184, 189–191, 232–233, 238–239, 280, 508
ON-state loss, 374, 482, 523
ON-state voltage, 11, 91, 256, 375, 518
ON-state voltage drop, 86, 240, 318, 421, 499, 501, 516, 542–544, 534–535,
ON-time, 551
Open-base transistor breakdown, 352
Open-loop control, 599–600
Operating states, thyristor, 189, 191
Operational modes of
  DG-IEGT, 524
  IGBT, 46
Optical lithography, 435–436
Orientation, 40, 113, 351, 412, 429–430, 457
Oscillations, gate voltage, 80
Oscillations, in IGBT module, 483, 485–486
Oxidation, thermal, 13, 429–431, 444, 447, 453–457
Oxide charge, 113, 118
Output capacitance, 18, 132
Output characteristics of
  BJT, 184–185
  IGBT, 51, 52, 229, 588
  MOSFET, 102
Output resistance, 122, 130–132
Overcurrent, 26, 480
Overcurrent protection, IGBT power module, 481
Overdriving, 185
Overlap capacitances, MOSFET, 130–132, 138
Overtemperature, 480
Overtemperature protection, IGBT power module, 481
Overvoltage, 26, 480
Oxidation, 13, 419, 429–431, 444, 447, 453–457
Oxide charge, 137, 441

Packages, 22, 83, 156–157, 443
Packaging of IGBT, 22, 83, 442–443
Parabolic oxidation law, 429, 457
Parabolic rate coefficient, 430, 457
Parallel and coupled PIN diode-PNP transistor model of carrier distribution in the ON-state of trench IGBT, 499–502

Parallel configuration, 5
Parallel effect, 502
Paralleling of IGBTs, 76, 83, 91, 465–470
Parallel plane junction, 82, 361, 386, 388, 393
  comparison of breakdown voltage with cylindrical junction, 406–410
Parameter extraction procedure for SPICE IGBT model, 587
Parameters section, SABER template, 590
Parasitic bipolar transistor in MOSFET, 8, 107, 136
Parasitic components, IGBT module, 482–484
Parasitic thyristor in IGBT, latch-up of, 303, 371
Partial discharge (PD), 489
  spectroscopy, 490
Passivation, 156
  of trench sidewall, 439
Pattern generator, electron-beam, 435
PCB. *See* printed circuit board
P-channel IGBT, 45
P-channel MOSFET, 119, 135, 509–510
Peak current, 187–188
Peak electric field, 163, 392
Peak power, 187–188
Peak-to- peak ripple current, 549, 551, 556
Peak-to- peak ripple voltage, 550
PECVD. *See* Plasma-enhanced CVD
PEMS. *See* Porcelain enameled metal substrate
Pencil coil, 602
Permalloy, 486
Permittivity, 161, 163, 249
Phase angle control, 570
Phase control thyristor, 2
Phase delay, 568
Phase locked loop (PLL) control, 598
Photolithography, 13, 106, 412–419, 435–436, 511, 518
Photoresist, 435
Physical models, DESSIS-ISE, 365–368
Physics-based IGBT circuit models, 589
Physics, IGBT, 229
PICASSO, 365
PICs. *See* power integrated circuits
Piecewise linear function, 589
Piling of impurities, 430
PIMS. *See* Polymer insulated metal substrate
Pinch-off, 120
P-I-N diode/rectifier, 16, 155, 172–176, 178, 193, 229–234, 376, 390–392, 502, 505, 518–519, 534
P-I-N rectifier-DMOSFET model of IGBT, 230–234, 241, 243, 245, 502–503
PISCES, 364
Planarization, 426
Plasma, current, 192
Plasma-enhanced CVD (PECVD), 420, 436–437
Plasma etching, 438
Plasma of holes, 50
Plastic package, 443
Plastic transfer molding, 475
PMOS transistor. *See* P-channel MOSFET
P-N junction, 155, 159, 167–170, 237, 403
  diode, 155–156, 225
  isolation, 475–478
PNP transistor, 14, 66, 71, 180, 183, 190, 225, 241, 249–250, 252, 254, 256, 258, 290, 306–307, 310, 312, 317, 319, 352, 376, 398, 403, 502–503, 518–519, 522–523, 525
Point-on-wave control, 570
Poisson's equation, 79, 109, 146, 357–359, 365, 367, 391, 404
Polyimide, 472–473
Polymer insulated metal substrate (PIMS), 472
Polymorphs, 350
Polynomial generator, 588
Polysilicon, 362, 418, 421, 426, 436–437
Polysilicon bars, 334
Polysilicon, heavily doped, 437
Polysilicon gate, 12, 13, 87, 114, 118, 335, 448, 511
  length, 125
Polysilicon window, 329, 331, 333–334
Polytypes, SiC, 350, 525
Porcelain enameled metal substrate (PEMS), 472
Positive photoresist, 435
Postmolded plastic package, 443
Potential, 159, 173
Potential difference, 142, 161
Potential distribution, 109, 173
Potential drop, 176, 217, 236–237, 293, 504, 542
Potential energy, 142
Power blackout, 597
Power burnout, 597
Power cycling, 485
Power diode, 155–157, 171–172
Power dissipation, 58, 91, 134, 184, 272–273
  critical, for IGBT, 280
Power factor, 571–574
Power IGBT modules, 465–498
Power integrated circuits (PICs), 516
Power loss, 20, 69, 133
Power module package, 22
  flat package, 484–486
  parasitic components, 482–484

Power module technologies, 470–475
Power MOSFET, 2–5, 103–107
Predeposition, 431–433, 444, 458–461
Press pack, 22, 488, 491
Pressure contact, 485
Printed circuit board, 472
Process and assembly sequence of IGBT fabrication, 448–449
Process design and fabrication technology of IGBT, 411–463
Process integration, 443–450
Process simulation, 443–450
Projected range, ion implantation, 433–434
Projected straggle, ion implantation, 433–434
Projection printing, 435
Proportional drive, 185
Protection of IGBT, 26, 466, 482, 479–482
Proton implantation, 61
Proton irradiation, 11, 422, 441
Proximity printing, 435
Pseudo-Darlington configuration, 14, 66
PT-IGBT. *See* Punch-through IGBT
P-type material, 157
Pulsed collector current, 91
Pulse-mode operation, 134
Pulse-width modulation (PWM), 27, 72, 93, 94, 545, 558, 565–568, 586
Punchthrough IGBT, 38, 39, 72, 74, 84, 177, 412, 421, 428
Punchthrough voltage, 355, 357, 372, 388–390, 406
Push-pull converter, 594
PWM. *See* Pulse width modulation
Pyrogenic oxidation, 414

QFC. *See* Fixed oxide charge
QRDCLI. *See* Quasi-resonant DC link inverter
Q_{rr}. *See* Reverse recovery charge
Q_{SS}. *See* Surface state charge
Quadratic equation, 457, 541
Quasi-equilibrium approximation, 254, 290
Quasi-Fermi level, 120
Quasi-Fermi potential, 368, 392
Quasi-neutral base region, 258
Quasi-neutral base width, 252, 258
Quasi-neutrality condition, 357
Quasi-resonant converter, 577
Quasi-resonant DC link inverter (QRDCLI), 585–586
Quasi-static (QS) approximation, 258
Quasi-stationary command, 365
Quasi-three-dimensional simulation, 392–393

Radiation damage, 441
Radiation damage coefficient, 67
Radiation effects, 67, 85, 421
Random thermal motion, 202
Raster scanning, 436
Ratings of IGBT, 90
Ratio of dynamic and static latching currents, 311–312
Rate equation for heat conduction, 278
RBSOA. *See* reverse-bias safe operating area
RDCLI. *See* Resonant DC link inverter
Reactive ion etching (RIE), 87, 106, 422, 512
Reactive plasma, 437
Reactive plasma etching, 437–439
Rechargeable battery unit, 597
Recombination, 64, 173, 177–178, 196, 259,
  Auger, 176–177
  lifetime, 170
  rate, 174, 206–207
  Shockley-Read-Hall, 367
Recombination-generation current, 182
Recovery diode, 104
Rectangular trench, 422
Rectifier unit, 597
Redistribution capacitance, 263
Reduced surface field technique, 395–398, 478
Reference level, 141
Reflow, 449
Refractory gate, 106
Regenerative braking, 557
Regenerative feedback, 14, 190
Reliabilty tests of IGBT, 80
Repetitive peak current, 155
Repetitive peak voltage, 155
Reset winding, 592–593
Residual base charge, 259
Resistive load turn-off in DC circuit, 308–309
Resistivity, 202, 351, 355–356, 440–441, 473
Resistor-inductor circuit, 171
Resistor-inductor load switching, 262
Resonant buck converter, 580
Resonant converter, 577
Resonant DC link inverter (RDCLI) 584–585, 587
Resonant link inverter, 583
Resonant pole inverter, 583–584
RESURF. *See* REduced SURface Field
RESURF concept for breakdown voltage enhancement, lateral IGBT, 395–398
Reticle, 435
Reverse-biased safe operating area (RBSOA), 79, 187, 334
Reverse blocking capability, 172, 509
  lateral IGBT with, 509
Reverse blocking mode of IGBT, 46
Reverse blocking state, thyristor, 191, 196

Reverse leakage current, 51, 53, 119, 160, 170
Reverse recovery, 304
  characteristics, 8, 171
  charge, 4, 8
  time, 4, 8, 155, 172
  transient, 196
Reverse saturation current, 136, 160, 190, 192, 194
Reverse transfer (Miller) capacitance, 93
Reverse voltage blocking capability, 172
Rice cooker, 604
RIE. *See* Reactive ion etching
Rise time, 56, 70, 91, 132, 161, 187, 195, 589
Rugged IGBT module, 482

SABER model, IGBT, 590–592
SABER package, 17
SABER template, 590
Safe operating area (SOA), 77, 79, 87, 107, 134, 188, 398, 421, 479, 525
SA-TIGBT. *See* Self-aligned trench IGBT
Saturated collector-emitter current, 244, 246
Saturated drain-source current, 122
Saturated drift velocity, 82, 276
Saturation current, 86, 170, 254
Saturation region, 51, 78, 184, 187–188, 197–198, 246
Saturation region MOSFET transconductance parameter (K_{psat}), 277
Saturation velocity, electron, 350
Saturation voltage, 91, 92, 184, 421, 588, 602
Scattering oxide, 414–419, 422, 448–449
Schwarz-Christoffel transformation, 239
SC-IGBT. *See* Silicon carbide IGBT
SCR. *See* Silicon-controlled rectifier
Screening length, 110
Second breakdown, 8, 78, 106, 134, 136, 187, 189
  bipolar, 136
  MOS, 136
Seed crystal, 349
Segmented collector IGBT (SC-IGBT), 37, 515
Segment length, 319–322
Segregation coefficient, 430
  effective, 430
  equilibrium, 430
Segregation of impurities, acceptor and donor, 430
Self-aligned sidewall-implanted N^+-emitter lateral IGBT (SI-LIGBT), 512–513
Self-aligned trench IGBT (SA-TIGBT), 505
Self-aligned VDMOS technology, 478
Self-clamped IGBT, 89
Self-clamped inductive switching (SCIS), 601–602
Self-clamping, 520
Self-isolation, 475–477
Semiconductor materials, properties, 350
Semiconductor potential, bulk, 152
Semiconductor selection for IGBT, 349–361
Semidiscrete heat equation, 297
Series configuration, 5
Shallow junction, 440
Shallow P^+ diffusion incorporation in IGBT cell, 322–324
Sheet charge, 116
Sheet resistance of P-base, 307, 317
Sheet resistivity, 330, 337
Shielding length, 110
Shockley equation, 170
Shockley-Read-Hall recombination, 367
Short-circuit current, 480
Short-circuit failure, 39, 482
Short-circuit protection, IGBT power module, 481
Short-circuit test, 81
Short-circuit withstand time, 93
Short-circuited transistor, 194–195
Shorted collector IGBT, 37
Shorted emitter construction, 193–194
Shorting resistor, 194
Shunting resistance, 319–320, 330–332, 335, 339, 341
SiC IGBT, 525
  4H-SiC IGBT, 350, 526
  6H-SiC IGBT, 350, 525
Sidewall implanted emitter, 512
Silane, 428, 437
Silicon carbide, 1,
  IGBT, 525–527
Silicon-controlled rectifier (SCR), 2, 23
Silicon dioxide, CVD, 437–438
Silicon dioxide, thermal, 429–431
Silicone, 443
Silicone gel, 493
Silicon nitride, CVD, 420, 422, 427, 437–438
Silicon-on-insulator (SOI) substrate, 476, 516–517
Silicon tetrachloride, 428
SI-LIGBT. *See* Self-aligned sidewall-implanted N^+-emitter lateral IGBT
Simulation, 229, 362–364, 368–372
Simulation hierarchy, 362–364
Single phase full-bridge inverter, 561–565
Single phase half-bridge inverter, 558–561
Single pulse width modulation, 567

Sinker, N^+, P^+, 511–512
Sintering, aluminum, 420, 427
Sinusoidal pulse width modulation (SPWM), 558, 566–567
SITh. *See* Static induction Thyristor
Slow recovery diode, 8, 155
Slow thyristor, 189
Smart IGBT module, 492
Smart power, 10, 466, 491, 515
SMPS. *See* Switch mode power supply
Snappy diode, 60
Snubber capacitor, 73
Snubber protection circuit, 229
SOA. *See* Safe operating area
Soft recovery diode, 479
Soft solder, 475
Soft switching, 72, 582–587
  DC-DC converter, 576–582
Solder, 440, 471, 475, 487–488, 490, 493–494
Solder fatigue, 475, 488
Solid angle, 239
Solid solubility, 431
Source, 101, 197–198
Source metallization capacitance, 258
Space-charge density, 146
Space-charge generation lifetime, 161, 170–171
Space-charge layer width, 145
Space-charge neutrality condition, 288
Space-charge region, 159, 169, 404
Spacing optimization of field rings, 391–392
Specific accumulation layer resistance, 148–149
Specific capacitance, 112
Specific channel resistance, 127, 129, 147–148
Specific drift region resistance, 149
Specific ON-resistance, 125, 127
Specific resistance, 125, 151–152
Specific transconductance, 121
SPICE, 18
  models of IGBT, 18, 19, 587
Spreading angle, 150
Spreading resistance, 125, 127, 151
Spreading velocity, plasma, 193
Sputter-etching, 437
SPWM. *See* Sinusoidal pulse width modulation
Square array, 138
Square cell, 138, 333, 371
Square wave AC, 604
Squirrel-cage induction motor, 599
Stability, 80
Static characteristics
  BJT, 180
  IGBT, 50–53, 512
  JFET, 197
  thyristor, 189
Static charge, 26
Static induction Thyristor (SITh), SiC, 526
Static latching, 303, 306–308, 342
Static model, 589
Static properties, 240
Stationary transport, 365
Steady-state analysis, 252
Steady-state forward conduction, 251, 372
Steady-state latching current, 307
Step junction, 353
Storage time, 187
Storage junction temperature ($T_{j,stg}$), 91
Stripe cell, 328–330
Strong inversion, 111–112, 114, 137, 146–147
Stud-type package, 156–157
Substrate resistance, 125
Substrate terminal, MOSFET, 102, 103, 131
Substrate thickness, SOI, 516
Surface breakdown, premature, 395
Surface charge, 110–111, 146–147, 393–394
Surface electric field control, 156
Surface ion implanted edge termination, 394–395
Surface leakage current, 182
Surface potential, 110–111, 145
Surface state charge Q_{ss}, 112, 351, 362, 411
Surface states, 351, 429
Surge current, 155
Sustaining voltage, 187, 189
Switchback, 52
Switching behavior, 172, 256
Switching circuit, 185
Switching energies, 92
Switching losses, 2, 9, 11, 40, 479, 482, 499, 520, 582
Switching speed, 398, 420–421
Switching time, 92, 130, 133, 185–186, 589
Switching waveforms, thyristor, 196
Switch mode power supply (SMPS), 4, 6, 592–596
Switch, power transistor, 184–185
Symbols, circuit diagram, of IGBT, 23
Symmetrical IGBT, 38, 40, 43
  latching of, with inductive load, 315–317
  latching of, with resistive load, 308–313
Synchronous motor, 600
System variables, 590

Tachometer, 600
Tail bump, 74

Tail plateau, 74
Tail, turn-off, 71, 72, 421
Taylor series, 203–204, 206
TCIGBT. *See* Trench clustered insulated gate bipolar transistor
TCR. *See* Temperature coefficient of resistance
Technological simulator, 447
Temperature coefficient of current, 83
Temperature coefficient of resistance, 11
Temperature dependency of parameters used in electrothermal simulations, 276–277
Temperature effect, 78, 83, 137, 171, 247, 373–377, 589
Temperature sensing, 479, 481
Template header, SABER template, 590
Terminal currents, transistor, 220
Terraced gate structure, 380–382
Thermal adhesive, 473
Thermal analysis, 246
Thermal behavior, 78, 83, 136
Thermal capacitance, 275
Thermal characteristics, 362
Thermal conductivity, 350, 476, 490–491
Thermal cycling, 488
Thermal destruction, 280
Thermal diffusion, 13, 444, 447
  cycles, 431–433
Thermal and electrical quantities, relationship between, 274
Thermal equilibrium, 109, 141, 159, 168, 176, 207, 217, 225
Thermal equivalent circuit, 278–279
Thermal expansion, coefficient of, 488, 490–491
Thermal fuse, 27
Thermal grease, 473, 490
Thermal impedance, 85, 274
Thermal interface pad, 473
Thermal network, 26, 274, 299
Thermal oxidation, 13, 418, 429–431, 444, 447, 453–457
Thermal resistance, 84, 495
Thermal stability, 39
Thermal tape, 473
Thermal velocity, 202
Thermal voltage, 137, 161, 168
Thermally assisted carrier multiplication, 81
Thermodynamic model, DESSIS, 365
Thermodynamic simulation, 229
Thermodynamic voltage, 276
Thermomechanical fatigue, 488
Thin film 473
Three-phase full-bridge inverter, 22, 479, 491–492, 568
Threshold voltage, 54, 76, 86, 92, 107, 111, 112, 114, 118, 129, 134–137, 247, 277, 326-327, 362, 364, 375, 440, 520
  variation with temperature, 137, 247
Thyristor, 2, 155–156, 188–189, 510, 513
  turn-off time, 195–196
  turn-on time, 196
TIGBT. *See* Trench-gate IGBT
TIG welding, 604
Time constant, 440
Titanium, 420
Topological designs, 138
TPIGBT. *See* Trench planar insulated gate bipolar transistor
t-plane, 239
Transconductance, 9, 54, 55, 81, 92–93, 119, 121–122, 132, 243–244, 246, 278, 362, 364, 368-372, 375, 381–383, 525, 588
Transfer characteristics, 53–55, 68, 376
Transient, 27, 172, 196, 258–259, 263–264, 582
Transient base stored charge Q_{base} (t), 500, 531–532
Transient characteristics, 364, 376, 380
Transient command, 365
Transient operation, 251–252
Transistor, bipolar, 2, 9, 104, 106, 156, 184, 220, 229
  terminal currents, 220
Transistor, MOSFET, 2–5, 9, 103–107, 136
Transmission line, 274
Transparent collector structure, 502
Transport equations, 236
Transverse electric field, 119
Transverse field transconductance factor (θ), 277
Trap, 182, 441
Trapezoidal field distribution, 42
Trapped positive hole charge, 134–135
Trench, 422, 425–426, 438–439
Trench cell, 240
Trench clustered insulated gate bipolar transistor (TCIGBT), 522–523
Trench formation, 438
Trench-gate IGBT (TIGBT), 86, 239, 340, 438, 527, 604
  fabrication, 422–427
  latching current density, 340–341
  parallel and coupled PIN diode-PNP transistor model of carrier distribution in the ON-state, 502–505
Trench-gate MOSFET, 138

Trench planar insulated gate bipolar transistor (TPIGBT), 518–520
Trench width, 126–127
Triangular field distribution, 42
Trichlorosilane, 428
Trigger angle advance, 93
Trigger angle delay, 93
Triggering angle, 93, 570
Trilayer metallization, 419–420, 427, 440
Triple diffused transistor, 179–180
Turn-off behavior of NPT-IGBT, non-quasi-static analytical models for, 266
Turn-off, BJT, 186
Turn-off energy loss, 91
Turn-off, IGBT, 64, 91
  with a resistive load, 67
  with an inductive load, 69
Turn-off loss, 39, 421, 523
Turn-off time, 11, 15, 21, 64, 66, 70, 132, 196, 376, 421, 441, 495
Turn-off transient, IGBT, 259
Turn-off voltage, variation with time, 265
Turn-off waveforms, 263
Turn-on, BJT, 186
Turn-on, IGBT, 55, 91
  with a resistive load, 56
  with an inductive load, 59
Turn-on time, 39, 56, 132, 195–196, 198, 495
Two-dimensional current flow, 238
Two-dimensional effects, 280
Two-dimensional model of carrier distribution, 238–239
Two-side gate, 505, 507
Two-transistor equivalent circuit, thyristor, 192

U-groove MOSFET, 87, 104–106, 123, 126, 128, 130
  model, 126
U-groove MOS-IGBT, 87, 340
Ultrasonic wire bonding, 442
Ultrathin substrate, 516
UMOSFET. *See* U-groove MOSFET
UMOS-IGBT. *See* U-groove MOS-IGBT
Undepleted base width in IGBT, 42, 44, 79, 165–166, 249, 252, 310, 352, 357
Undervoltage lockout, IGBT power module, 481
Uninterruptible power supply (UPS), 2, 6, 7, 491, 596–597, 604
Unipolar
  conduction, 3
  device, 104
Unit cell, 12, 13, 14, 119, 129–130, 138, 147–151, 237, 244, 246, 330, 338, 349, 368–371
Unit process steps, IGBT, 428–443
Unsaturated bond, 113
UPS. *See* Uninterruptible power supply

Vacuum level, 108
Valence band, 109, 141, 146, 441
Values section, SABER template, 590
Vapor phase epitaxy, 428
Varistor, 26
V_{BR} equation, 403–404
VCVS. *See* Voltage-controlled voltage source
VDMOSFET. *See* Vertical double-diffused MOSFET
VDMOSFET IGBT fabrication, 412–422
  flow chart, 414–415
Vector control technique, 600
Vector scanning, 436
Vertical double-diffused MOSFET (VDMOSFET), 2, 103–106, 256
Vertical IGBT (VIGBT), 35
VFI. *See* Voltage-fed inverter
V-groove MOSFET, 2, 103–104,
VIGBT. *See* Vertical IGBT
Virtual base contact, 241–242
VMOS-IGBT, 87
Void layer, 441
Void, solder, 440
Voltage amplification, 123
Voltage-boundary condition, 81
Voltage-controlled current source (VCCS), 18
Voltage-controlled oscillator, 600
Voltage-controlled resistor, 198
Voltage drop, 219, 230, 238
  accumulation region, 247–248
  channel, 247
  drain-source, 255
  emitter-base, 254
  JFET region, 248
  middle N^- base, 175, 177–178, 219
  N^- drift region, 237
  ohmic, 175
Voltage-fed inverter (VFI), 558
Voltage gate drive circuit of IGBT, 24, 25
Voltage rating, 155
Voltage sharing, 469
Voltage source, 229, 588
Voltage-source inverter (VSI), 558, 599–600

Voltage-to-frequency converter, 600
Volt-ampere characteristic, 170
Volumetric heat capacity, 271
VSI. *See* Voltage-source inverter

Water vapor, 429
Weak inversion, 111
Webster's effect, 182
Wedge bond, 475, 487
Welding, 491, 602
Well,
  N-well, 520–523
  P-well, 520–523
Wet oxidation, silicon, 429–431
Wetting, solder, 420, 440
Width of the channel, 119
Width of the unit cell, 150
Wire bond, 475, 487
Wire bonding, 420, 442, 494
Work function, 107, 112, 116, 137, 326
Wound rotor induction motor, 600

X-ray lithography, 435–436
X-ray tubes, 6

Young's modulus, 490–491

ZCS-QRC. *See* Zero current switching-quasi resonant cell
ZCS resonant buck converter, 580
Zener diode, 23
Zero-bias depletion region width (W_0), 150, 248
Zero-current switching (ZCS), 72, 74, 484, 575–576, 578, 580
Zero current switching-quasi resonant cell (ZCS-QRC), 577–578
Zero-voltage switching (ZVS), 72, 75, 484, 575–576, 578, 583, 586
Zero voltage switching-quasi resonant cell (ZVS-QRC), 577–578
ZVS buck converter, 578
ZVS-QRC. *See* Zero voltage switching-quasi resonant cell